

CLASIFICACIÓN DE EVENTOS SÍSMICOS EMPLEANDO PROCESOS GAUSSIANOS

Event Seismic Classification using Gaussian Processes

RESUMEN

La clasificación de señales sísmicas es de crucial importancia para el descubrimiento de posibles interacciones entre movimientos telúricos volcánicos y procesos volcánicos *per se*. En este artículo, se presenta la aplicación de procesos gaussianos para la clasificación de eventos sísmicos registrados en el volcán Nevado del Ruíz. Las señales se caracterizan usando los coeficientes de un modelo autoregresivo, empleado para estimar la densidad espectral de potencia. La función de distribución predictiva para la clasificación se aproxima mediante el método de Laplace. El desempeño obtenido es mayor que el de una red neuronal artificial, clasificador utilizado tradicionalmente para resolver esta tarea.

PALABRAS CLAVES: eventos sísmicos, clasificación, procesos gaussianos, Nevado del Ruíz.

ABSTRACT

Seismic signals classification is important by itself in order to discover factual interactions between volcanic earthquakes and volcanic processes. In this paper, it is presented the application of Gaussian processes for seismic events classification registered at Nevado del Ruiz volcano. Feature extraction is accomplished using the coefficients of an autoregressive model, employed for the estimation of the power spectral density. The predictive distribution for classification is approximated using the Laplace method. Obtained performance is higher than the one obtained with an artificial neural network, the state of the art classifier for this kind of task.

KEYWORDS: seismic events, classification, Gaussian processes, Nevado del Ruiz.

1. INTRODUCCIÓN

El volcán nevado del Ruiz está cubierto por una gran cantidad de nieve y hielo, formando un glaciar cuyo volumen es de aproximadamente 1300 millones de metros cúbicos. Posee tres cráteres: Arenas (que permanece activo), y dos cráteres parásitos, Olleta y Piraña. Debido a que la actividad sísmica del volcán ha sido grabada en formato digital por el Observatorio Vulcanológico y Sismológico de Manizales (OVSM), se ha producido una gran cantidad de datos en las redes de monitoreo, por lo cual resulta ser apropiado la aplicación de técnicas automáticas de clasificación.

La clasificación de señales sísmicas es de crucial importancia para el descubrimiento de posibles interacciones entre movimientos telúricos volcánicos y procesos volcánicos *per se*. En este trabajo, se consideran dos clases de señales sísmicas originadas en el nevado del Ruiz, eventos vulcano-tectónicos (VT) y de largo período (LP). Si bien a grandes rasgos, estas son las dos clases más generales de eventos sísmicos, los sismólogos establecen independientemente sus propios subtipos de

acuerdo a aspectos particulares de cada volcán [1]. Actualmente, los expertos del OVSM clasifican los tipos de eventos sísmicos registrados mediante inspección visual, lo cual sugiere una gran cantidad de carga laboral para los analistas de señales sísmicas. En consecuencia, una herramienta de clasificación automática reduciría tal cantidad de trabajo, produciendo además, una clasificación confiable y efectiva, y eliminando los errores asociados a evaluaciones tediosas y al cambio de personal.

Entre las aplicaciones de reconocimiento de patrones a señales sísmicas, entre las más recientes sobresalen: clasificación automática de señales sísmicas del volcán Monte Vesubio (Italia) [2] y clasificación automática de eventos sísmicos del volcán Colinas Sufrière (Monserate) [3]. Ambas, proponen como clasificador una red neuronal artificial (ANN). En [2], un perceptrón multicapa (MLP) se emplea para distinguir entre eventos VT y señales transitorias producidas por otras fuentes como explosiones bajo el agua, ráfagas explosivas en minas y rayos; mediante características espectrales y parámetros de amplitud. En [3], se utiliza una ANN para

MAURICIO ALVAREZ

Ingeniero Electrónico, M.Sc.

Profesor Auxiliar

Universidad Tecnológica de Pereira

malvarez@utp.edu.co

RICARDO HENAO

Ingeniero Electrónico, M.Sc.

Profesor Auxiliar

Universidad Tecnológica de Pereira

rhenao@utp.edu.co

EDISON DUQUE

Ingeniero Electrónico, M.Sc. (c)

Profesor Asistente

Universidad Tecnológica de Pereira

eduke@utp.edu.co

clasificar cinco clases fundamentales de señales, VT, regionales (RE), LP, híbridos (HB) y avalanchas (AV); caracterizando las señales con funciones de autocorrelación, momentos estadísticos de orden superior y razones de amplitud; reportando una tasa de error de 30%, la cual fue reducida a 20% luego de una revisión manual de las etiquetas inicialmente contenidas en la base de datos. Convencionalmente, en el contexto de sismología volcánica, las redes neuronales han sido preferidas en lugar de otras técnicas clásicas de reconocimiento de patrones. La popularidad de los modelos de redes neuronales para resolver problemas de reconocimiento de patrones se debe principalmente a la poca dependencia que presentan con respecto al conocimiento específico del dominio de los datos y a la disponibilidad de algoritmos de entrenamiento eficientes [4]. Además, las redes neuronales permiten utilizar funciones base *adaptativas*, a diferencia de los modelos lineales convencionales. Dichas funciones *adaptativas* o unidades ocultas permiten capturar características subyacentes en los datos que pueden ser útiles para modelar el problema que se está tratando. Sin embargo, la adaptabilidad produce una gran cantidad de inconvenientes prácticos, por ejemplo, sobre entrenamiento crítico, cuando el número de funciones base incrementa y demasiados parámetros libres para establecer la arquitectura y el algoritmo de entrenamiento de la red neuronal.

Desde el punto de vista de las redes neuronales, hay básicamente dos formas de mejorar sus capacidades de modelamiento, (i) incrementar la flexibilidad de las unidades ocultas, lo cual además del sobre entrenamiento, genera la necesidad de elegir entre una gran cantidad de parámetros que hacen posible tal flexibilidad y (ii) incrementar substancialmente el número de funciones base (tal vez a infinito numerable), que resulta en un problema mayor desde el punto de vista computacional, al tener que realizar todos los cálculos necesarios en un tiempo *acceptable*. El primer inconveniente se puede resolver parcialmente con técnicas de regularización. El segundo, de otro lado, se puede abordar con una técnica de aprendizaje de máquina denominada Procesos Gaussianos [5]. Los Procesos Gaussianos son una generalización de la distribución de probabilidad gaussiana, debido a que mientras una distribución de probabilidad describe variables aleatorias que son escalares o vectores, un proceso estocástico gaussiano controla las propiedades de dichas funciones, a partir de una cantidad finita de datos, produciendo una alta capacidad de generalización del modelo sin el inconveniente de hacer tender el número de funciones base explícitamente a infinito. Por lo tanto, una de los principales atractivos de los procesos gaussianos es que combina un concepto sofisticado y consistente con tratabilidad computacional. Además, resultan ser equivalentes matemáticamente a modelos como, los lineales bayesianos, los basados en splines y a las redes

neuronales con infinito numerable cantidad de unidades ocultas [5].

2. BASE DE DATOS

La base de datos nevado del Ruiz contiene señales seleccionadas de la colección de registros de la red de monitoreo de OVSM. Las estaciones de monitoreo de la red están localizadas en puntos estratégicos, por ejemplo, cerca al glaciar y a los cráteres. Para los experimentos se eligieron las señales registradas por dos estaciones en particular, Olleta (en el cráter) y Glaciar, debido a que de acuerdo con la experiencia del personal del OVSM, son el punto de referencia para eventos tanto volcánicos como tectónicos. La estaciones están localizadas a 4.08 Km y 1.8 Km del cráter activo, respectivamente. Las señales fueron digitalizadas con una frecuencia de muestreo de 100.16 Hz utilizando un convertidor digital análogo de 12 bits. La base de datos está compuesta por 483 señales Vulcano-Tectónicas (VT) y 580 señales Largo- Período (LP).

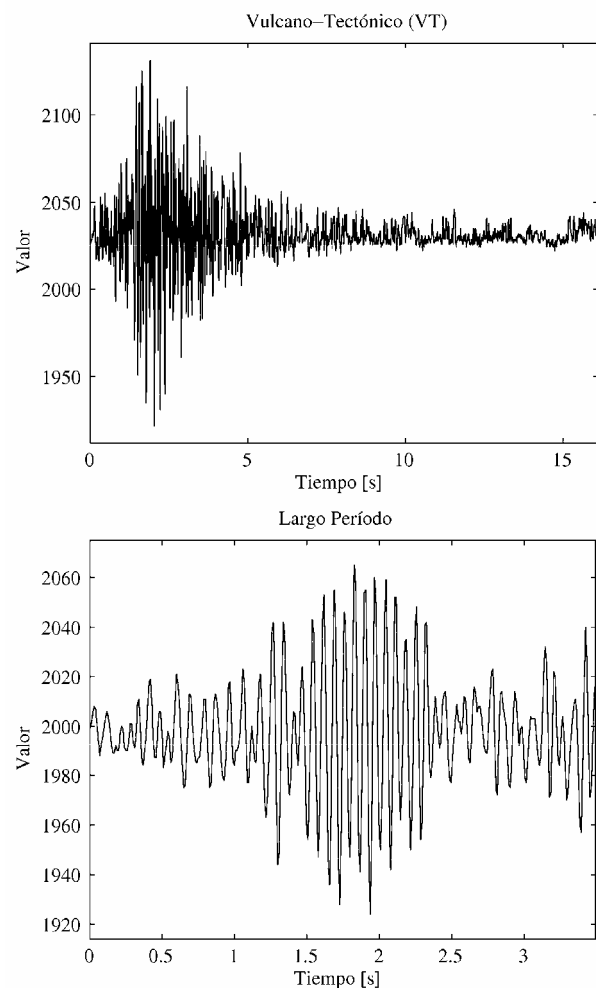


Figura 1. Formas de onda características de las señales de las clases VT y LP respectivamente

3. EXTRACCIÓN DE CARACTERÍSTICAS

El espectro de los registros se usa comúnmente para la clasificación y el monitoreo de la actividad sísmica. Debido a que las diferencias en el contenido espectral permiten la discriminación visual de los tipos diferentes de movimientos telúricos volcánicos [1], en este artículo se realiza la caracterización con estimación paramétrica de la densidad espectral de potencia usando un modelo autoregresivo. Los coeficientes del modelo se estiman mediante tres métodos diferentes: (i) Yule-Walker (YULEAR) (ii) método de la covarianza modificada (MCOV) y (iii) el método de Burg (PBURG). Antes de calcular el espectro, se remueve la componente DC de los registros y se normalizan con respecto al área de la densidad espectral estimada. Para el cálculo del espectro, se emplea una FFT de 128 puntos, que corresponde a la longitud del registro más corto. Mediante pruebas preliminares, se selecciona el orden del modelo autoregresivo y se escoge igual a 6, con base en los resultados del error de clasificación.

4. PROCESOS GAUSSIANOS

Considérese un modelo definido en términos de una combinación lineal de M funciones base dadas por los elementos del vector $\phi(\mathbf{x})$ de forma tal que

$$y(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) \quad (1.1)$$

donde \mathbf{x} es el vector de entrada y \mathbf{w} es un vector de pesos de dimensión M . En la práctica, se desea evaluar la función en un conjunto finito de valores de \mathbf{x} , por ejemplo en los puntos de entrenamiento $\mathbf{x}_1, \dots, \mathbf{x}_N$. Se desea modelar la probabilidad conjunta $y(\mathbf{x}_1), \dots, y(\mathbf{x}_N)$, que se denota por el vector \mathbf{y} con elementos $y_n = y(\mathbf{x}_n)$ para $n = 1, \dots, N$. De la ecuación (1.1), este vector está dado por

$$\mathbf{y} = \Phi \mathbf{w}$$

donde Φ es una matriz de diseño con elementos $\Phi_{nk} = \phi_k(\mathbf{x}_n)$. Asumiendo que $p(\mathbf{w}) = N(\mathbf{w} | \mathbf{0}, \alpha^{-1} \mathbf{I})$, la función de distribución sobre \mathbf{y} sigue igualmente una distribución gaussiana $p(\mathbf{y}) = N(\mathbf{y} | \mathbf{0}, \mathbf{K})$, donde \mathbf{K} está dada como [7]

$$\mathbf{K} = \frac{1}{\alpha} \Phi \Phi^T$$

La función de probabilidad anterior $p(\mathbf{y})$ es un caso particular de un proceso gaussiano. En forma general, un *proceso gaussiano* se define como una distribución de probabilidad sobre funciones $y(\mathbf{x})$ tal que el conjunto de

valores de $y(\mathbf{x})$ evaluados en un conjunto arbitrario de puntos $\mathbf{x}_1, \dots, \mathbf{x}_N$ tienen conjuntamente una función de distribución gaussiana.

Un punto importante de los procesos estocásticos gaussianos es que la distribución conjunta sobre N variables y_1, \dots, y_N se especifica completamente con estadística de segundo orden. Para muchas aplicaciones se asume que la media del proceso es igual a cero y la matriz de covarianza se especifica evaluando $y(\mathbf{x})$ para dos valores de \mathbf{x} , lo cual permite obtener

$$E[y(\mathbf{x}_n)y(\mathbf{x}_m)] = k(\mathbf{x}_n, \mathbf{x}_m)$$

donde $k(\mathbf{x}, \mathbf{x}')$ se conoce como la función kernel [8]. Una condición necesaria y suficiente para que la función $k(\mathbf{x}, \mathbf{x}')$ sea un kernel válido [9] es que la matriz de Gram \mathbf{K} , cuyos elementos están dados por $k(\mathbf{x}_n, \mathbf{x}_m)$, debe ser semidefinida positiva para todos los valores posibles de $\{\mathbf{x}_n\}$.

4.1 Procesos gaussianos para clasificación

En un enfoque probabilístico para clasificación, la idea es modelar las probabilidades a posteriori de una variable objetivo (la etiqueta de clase) para un nuevo vector de entrada, dado un conjunto de datos de entrenamiento. Considérese un problema de dos clases con una variable objetivo $t \in \{0, 1\}$. Cuando se emplea un proceso gaussiano para clasificación, primero se define un proceso gaussiano sobre la función $a(\mathbf{x})$, que luego se transforma usando una función logística sigmoideal $y = \sigma(a)$, donde

$$\sigma(a) = \frac{1}{1 + \exp(-a)}$$

Se obtiene entonces un proceso estocástico no gaussiano sobre las funciones $y(\mathbf{x})$ con $y \in \{0, 1\}$.

Formalmente, sea $\mathbf{x}_1, \dots, \mathbf{x}_N$ el conjunto de vectores de entrenamiento con sus correspondientes etiquetas de clase $\mathbf{t} = (t_1, \dots, t_N)^T$. Sea \mathbf{x}_{N+1} un vector de prueba con su correspondiente etiqueta t_{N+1} . El objetivo es determinar la función de distribución predictiva $p(t_{N+1} | \mathbf{t})$. Para lograrlo, se introduce un proceso gaussiano sobre el vector \mathbf{a}_{N+1} que tiene componentes $a(\mathbf{x}_1), \dots, a(\mathbf{x}_{N+1})$. A su turno, esto define un proceso no gaussiano sobre \mathbf{t}_{N+1} y, condicionándolo sobre el conjunto de entrenamiento \mathbf{t}_N , se obtiene la distribución predictiva que se requiere. El proceso gaussiano para \mathbf{a}_{N+1} toma la forma

$$p(\mathbf{a}_{N+1}) = \mathbf{N}(\mathbf{a}_{N+1} | \mathbf{0}, \mathbf{C}_{N+1})$$

donde la matriz de covarianza tiene elementos dados por

$$\mathbf{C}(\mathbf{x}_n, \mathbf{x}_m) = k(\mathbf{x}_n, \mathbf{x}_m) + \nu \delta_{nm}$$

siendo $k(\mathbf{x}_n, \mathbf{x}_m)$ una función kernel semidefinida positiva y la constante ν se introduce para asegurar que la matriz \mathbf{C} sea definida positiva. Una función kernel ampliamente utilizada es

$$k(\mathbf{x}_n, \mathbf{x}_m) = \theta_0 \exp\left\{-\frac{\theta_1}{2} \|\mathbf{x}_n - \mathbf{x}_m\|^2\right\} + \theta_2 + \theta_3 \mathbf{x}_n^T \mathbf{x}_m \quad (1.2)$$

Para el problema biclase es suficiente con predecir $p(t_{N+1}=1 | \mathbf{t}_N)$, dado que el valor de $p(t_{N+1}=0 | \mathbf{t}_N)$ se obtiene como $1 - p(t_{N+1}=1 | \mathbf{t}_N)$. La distribución predictiva requerida está dada por

$$p(t_{N+1}=1 | \mathbf{t}_N) = \int p(t_{N+1}=1 | a_{N+1}) p(a_{N+1} | \mathbf{t}_N) da_{N+1} \quad (1.3)$$

donde $p(t_{N+1}=1 | a_{N+1}) = \sigma(a_{N+1})$.

Esta integral es intratable y debe emplearse algún método para aproximarla. En este trabajo, se emplea el método de Laplace [10]. La idea es encontrar una aproximación gaussiana a la distribución posterior sobre a_{N+1} , que usando el teorema de Bayes, está dada por,

$$\begin{aligned} p(a_{N+1} | \mathbf{t}_N) &= \int p(a_{N+1}, \mathbf{a}_N | \mathbf{t}_N) d\mathbf{a}_N \quad (1.4) \\ &= \frac{1}{p(\mathbf{t}_N)} \int p(a_{N+1}, \mathbf{a}_N) p(\mathbf{t}_N | a_{N+1}, \mathbf{a}_N) d\mathbf{a}_N \\ &= \frac{1}{p(\mathbf{t}_N)} \int p(a_{N+1} | \mathbf{a}_N) p(\mathbf{a}_N) p(\mathbf{t}_N | \mathbf{a}_N) d\mathbf{a}_N \\ &= \int p(a_{N+1} | \mathbf{a}_N) p(\mathbf{a}_N | \mathbf{t}_N) d\mathbf{a}_N \end{aligned}$$

donde se ha usado $p(\mathbf{t}_N | a_{N+1}, \mathbf{a}_N) = p(\mathbf{t}_N | \mathbf{a}_N)$. Si se asume que la función $a(\mathbf{x})$ sigue la forma lineal de la ecuación (1.1), se puede demostrar [7] que la función de probabilidad $p(a_{N+1} | \mathbf{a}_N)$ tiene la forma

$$p(a_{N+1} | \mathbf{a}_N) = \mathbf{N}(a_{N+1} | \mathbf{k}^T \mathbf{C}_N^{-1} \mathbf{a}_N, c - \mathbf{k}^T \mathbf{C}_N^{-1} \mathbf{k}) \quad (1.5)$$

La integral (1.4) se puede evaluar encontrando una aproximación de Laplace para la distribución posterior $p(\mathbf{a}_N | \mathbf{t}_N)$ y usando el resultado estándar para la convolución de dos distribuciones gaussianas.

La distribución a-priori $p(\mathbf{a}_N)$ está dada por un proceso gaussiano de media cero con matriz de covarianza \mathbf{C}_N y la verosimilitud de los datos (asumiendo independencia estadística) está dada por

$$p(\mathbf{t}_N | \mathbf{a}_N) = \prod_{n=1}^N \sigma(a_n)^{t_n} (1 - \sigma(a_n))^{1-t_n} = \prod_{n=1}^N \exp(a_n t_n) \sigma(-a_n)$$

La aproximación de Laplace se obtiene haciendo una expansión en serie de Taylor para el logaritmo del término $p(\mathbf{a}_N | \mathbf{t}_N)$. Este logaritmo está dado como

$$\begin{aligned} \psi(\mathbf{a}_N) &= \ln p(\mathbf{a}_N) + \ln p(\mathbf{t}_N | \mathbf{a}_N) \\ &= -\frac{1}{2} \mathbf{a}_N^T \mathbf{C}_{N+1}^{-1} \mathbf{a}_N - \frac{N}{2} \ln(2\pi) - \frac{1}{2} \ln |\mathbf{C}_{N+1}| \quad (1.6) \\ &\quad + \mathbf{t}_N^T \mathbf{a}_N - \sum_{n=1}^N \ln(1 + \exp(a_n)) + \text{const} \end{aligned}$$

El modo de la distribución posterior se puede encontrar del gradiente de la expresión anterior. El gradiente está dado como

$$\nabla \psi(\mathbf{a}_N) = \mathbf{t}_N - \boldsymbol{\sigma}_N - \mathbf{C}_N^{-1} \mathbf{a}_N,$$

donde $\boldsymbol{\sigma}_N$ es un vector de elementos $\sigma(a_n)$. El valor de \mathbf{a}_N , que hace cero el gradiente, se encuentra utilizando un proceso recursivo conocido como el algoritmo de mínimos cuadrados recursivos reponderados (IRLS). La fórmula de actualización iterativa para \mathbf{a}_N está dada por

$$\mathbf{a}_N^{\text{nuevo}} = \mathbf{C}_N (\mathbf{I} + \mathbf{W}_N \mathbf{C}_N)^{-1} \{ \mathbf{t}_N - \boldsymbol{\sigma}_N + \mathbf{W}_N \mathbf{a}_N \},$$

donde \mathbf{W}_N es una matriz diagonal con elementos $\sigma(a_n)(1 - \sigma(a_n))$. En el modo, el gradiente $\nabla \psi(\mathbf{a}_N)$ se desvanece y la solución \mathbf{a}_N^* satisface,

$$\mathbf{a}_N^* = \mathbf{C}_N (\mathbf{t}_N - \boldsymbol{\sigma}_N),$$

Una vez se encuentra el modo \mathbf{a}_N^* del posterior, la matriz hessiana se determina como

$$\mathbf{H} = -\nabla \nabla \psi(\mathbf{a}_N) = \mathbf{W}_N + \mathbf{C}_N^{-1},$$

donde los elementos de \mathbf{W}_N se evalúan usando \mathbf{a}_N^* . Lo anterior define una distribución gaussiana que sirve como aproximación a la distribución posterior $p(\mathbf{a}_N | \mathbf{t}_N)$, como

$$q(\mathbf{a}_N) = \mathbf{N}(\mathbf{a}_N | \mathbf{a}_N^*, \mathbf{H}^{-1})$$

Combinando la expresión anterior con la expresión (1.5) y evaluando a continuación la integral (1.4), se tiene que

$$p(a_{N+1} | \mathbf{t}_N) = N(a_{N+1} | \mathbf{k}^T(\mathbf{t}_N - \boldsymbol{\sigma}_N), c - \mathbf{k}^T(\mathbf{W}_N^{-1} + \mathbf{C}_N)^{-1} \mathbf{k})$$

Debido a que $p(a_{N+1} | \mathbf{t}_N)$ sigue una distribución gaussiana, para evaluar la expresión (1.3) se puede emplear la aproximación

$$\int \sigma(a) N(a | \mu, \sigma^2) da \approx \sigma(\kappa(\sigma^2) \mu)$$

donde $\kappa(\sigma^2) = (1 + \pi\sigma^2/8)^{-1/2}$.

4.2 Determinación de los hiperparámetros

Una forma de determinar los parámetros $\boldsymbol{\theta}$ de la matriz de covarianza (ecuación (1.2)), consiste en maximizar la función de verosimilitud $p(\mathbf{t}_N | \boldsymbol{\theta})$, para lo cual, se necesitan expresiones para la función de verosimilitud logarítmica y su gradiente. La función de verosimilitud se define por

$$p(\mathbf{t}_N | \boldsymbol{\theta}) = \int p(\mathbf{t}_N | \mathbf{a}_N) p(\mathbf{a}_N | \boldsymbol{\theta}) d\mathbf{a}_N$$

De nuevo esta integral es intratable. Usando la aproximación de Laplace se obtiene la siguiente expresión para la verosimilitud logarítmica,

$$\ln p(\mathbf{t}_N | \boldsymbol{\theta}) = \psi(\mathbf{a}_N^*) - \frac{1}{2} \ln |\mathbf{W}_N^{-1} + \mathbf{C}_N| + \frac{N}{2} \ln(2\pi)$$

donde $\psi(\mathbf{a}_N^*) = \ln p(\mathbf{a}_N^* | \boldsymbol{\theta}) + \ln p(\mathbf{t}_N | \mathbf{a}_N^*)$. La expresión para las derivadas de $\ln p(\mathbf{t}_N | \boldsymbol{\theta})$ con respecto a $\boldsymbol{\theta}$, pueden encontrarse en [7].

5. RESULTADOS

Los experimentos fueron realizados para comparar el desempeño de los procesos gaussianos entrenados con las representaciones espectrales mencionadas anteriormente (MCOV, PBURG y YULEAR). En este caso, se utilizó la matriz de covarianza exponencial cuadrática (ecuación (1.2) haciendo θ_2 y θ_3 iguales a cero) y sus parámetros fueron encontrados automáticamente (ver sección 4.2) empleando gradiente conjugado como algoritmo de optimización. Los criterios de validación empleados fueron, validación cruzada de 5 particiones y curvas ROC (Receive Operator Characteristic Curve), en particular el área bajo la curva (AUC) [11]. En la tabla 2 se presentan los resultados obtenidos y en las figura 2 se muestran las curvas ROC obtenidas para cada una de las representaciones espectrales empleadas. Los resultados para MLP fueron extraídos de [12].

| Representación Espectral | Precisión | AUC |
|--------------------------|------------|-------|
| GP-MCOV | 88.06±3.06 | 0.939 |
| GP-PBURG | 87.57±2.24 | 0.939 |
| GP-YULEAR | 88.06±1.48 | 0.941 |
| MLP-YULEAR | 87.46±2.90 | -- |

Tabla 2. Resultados obtenidos con procesos gaussianos para las representaciones espectrales empleadas, MCOV, PBURG y PYULEAR

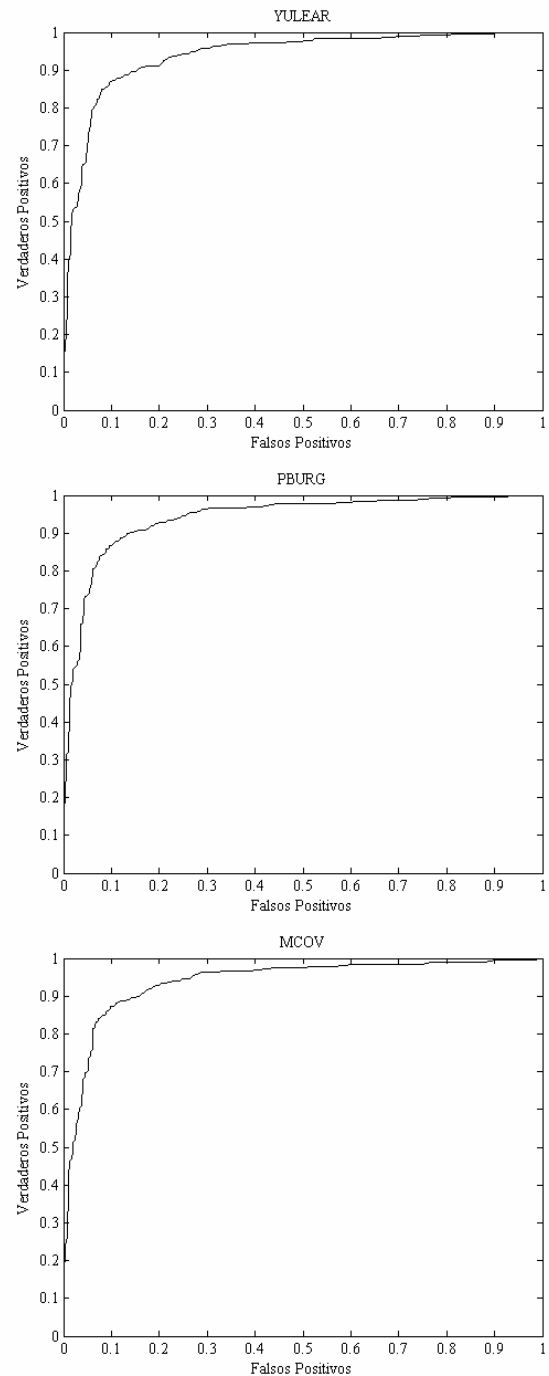


Figura 2. Curvas ROC para los GP entrenados sobre YULEAR, PBURG y MCOV respectivamente

De la tabla 2, se puede observar que los procesos gaussianos presentan un mejor desempeño en cuanto a procesión de clasificación con respecto a las redes neuronales de tipo MLP. En particular, el mejor resultado se obtiene con la representación basada en estimación espectral con el método de Yule-Walker, tanto en términos de precisión (y desviación estándar) como de área bajo la curva ROC.

6. CONCLUSIONES

En este artículo se ha mostrado que la utilización de procesos gaussianos para clasificación de eventos sísmicos permite obtener tasas de clasificación comparables a las alcanzadas mediante redes neuronales, sin el inconveniente de la selección de la topología de la red. Adicionalmente, se mostró que la representación espectral utilizando el método de Yule-Walker permite obtener tasas de clasificación y AUC superiores. El motivo de este resultado se encuentra en el hecho de que el proceso gaussiano es una generalización de la red neuronal que incorpora un número infinito de funciones base, haciendo que el modelo sea más flexible en la explicación de los datos.

La utilización de procesos gaussianos en lugar de MLP resulta ser ventajosa desde el punto de vista práctico debido a que la determinación de los parámetros libres del modelo se hace automáticamente dentro del contexto bayesiano. Además, dentro del mismo contexto, hace al clasificador menos sensible al fenómeno de sobre entrenamiento.

Si bien la formulación de procesos gaussianos es consistente desde el punto de vista teórico, la generalización para el caso de clasificadores con múltiples clases no es inmediata, como en el caso de redes neuronales, si se quisieran considerar múltiples tipos de sismos. Sería interesante explorar formas alternativas de caracterización que permitieran explotar de mejor manera la información contenida en las señales o, de otro lado, considerar la posibilidad de considerar registros tridimensionales.

7. AGRADECIMIENTOS

Este trabajo es financiado por el Centro de Investigación y Extensión (CIE) mediante el proyecto *Clasificación Automatizada de Eventos Sísmicos*, contrato 9-07-4.

8. BIBLIOGRAFÍA

- [1] V. Zobin. Introduction to Volcanic Seismology. Elsevier. 2003.
- [2] S. Scarpetta, F. Giudicepetro, E. Ezin, S. Petrosino, E. Pezzo, M. Martini y M. Marinaro. Automatic Classification of Seismic Signals at Mt. Vesuvius, Italy, using Neural Networks. Bulletin of the Seismology Society of America. Vol. 95(1). 2005. pp. 185-196.
- [3] H. Langer, S. Falsaperla, T. Powell y G. Thompson. Automatic Classification and a-Posteriori Analysis of Seismic Events Identification at Soufrière Hills Volcano, Monserrat. Journal of Volcanology and Geothermal Research. Vo. 153. 2006. pp. 1-10.
- [4] A.K. Jain, R.P.W. Duin y J. Mao. Statistical Pattern Recognition: A Review. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 22(1). 2000. pp. 4-37.
- [5] C.E. Rasmussen y C.K.I. Williams. Gaussian Processes for Machine Learning. The MIT Press. 2006.
- [6] D. Manolakis, V.K. Ingle, S.M. Kogon. Statistical and Adaptive Signal Processing: Spectral Estimation, Signal Modeling, Adaptive Filtering and Array Processing. Artech House Publishers. 2005.
- [7] C. Bishop. Pattern Recognition and Machine Learning. Springer. 2006.
- [8] B. Schölkopf, A. Smola. Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond. The MIT Press. 2001.
- [9] J. Shawe-Taylor, N. Cristianini. Kernel Methods for Pattern Analysis. Cambridge University Press. 2004.
- [10] D. MacKay. Information Theory, Inference & Learning Algorithms. Cambridge University Press. 2002.
- [11] T. Fawcett. An Introduction to ROC Analysis. Pattern Recognition Letters. Vol. 27. 2006. pp. 861-874.
- [12] M. Orozco-Alzate, M. Alvarez, R. Henao, C.G. Castellanos y R.P.W. Duin. Dissimilarity-Based Classification of seismic Signals at Nevado del Ruiz Volcano. Submitted to Journal of Volcanology and Geothermal Research. 2006.