

COMPARACIÓN ENTRE ENFOQUES GENERATIVOS Y DISCRIMINATIVOS PARA LA CLASIFICACIÓN DE EVENTOS SÍSMICOS

Comparison between generative and discriminative approaches for classification of seismic events

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 compara el desempeño de dos clasificadores que se enmarcan en metodologías diferentes de clasificación: modelos de mezcla de gaussianas, dentro de la estrategia de clasificación usando modelos generativos y máquinas de vectores de soporte, en la estrategia de clasificación que usa funciones discriminativas. Igualmente, se presentan diferentes combinaciones de estos dos clasificadores. Los resultados muestran que la estrategia basada en función discriminativa entrega porcentajes de acierto superiores.

PALABRAS CLAVES: eventos sísmicos, clasificación, modelo generativo, función discriminativa, 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, the performance of two classifiers under different classification strategies is compared: Gaussian mixture models, under the generative model classification approach and support vector machines, below the discriminative function classification approach. Likewise, different combinations of both classifiers are presented. Results using the discriminative function approach are superior.

KEYWORDS: seismic events, classification, generative model, discriminative function, 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 tres clases de señales sísmicas originadas en el nevado del Ruiz, eventos volcano-tectónicos (VT), de largo período (LP) y eventos relacionados con hielo (IC). Si bien a grandes rasgos, éstas son las tres 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 [9].

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 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

BONIE J. RESTREPO

Ingeniero Electricista.
Profesor Catedrático
Universidad Tecnológica de Pereira
bonie@ohm.utp.edu.co

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

30%, la cual fue reducida a 20% luego de una revisión manual de las etiquetas inicialmente contenidas en la base de datos.

En este artículo se compara el desempeño de dos metodologías de clasificación diferentes: una de ellas basada en modelos generativos y la otra en funciones discriminativas. Los modelos generativos pretenden dar cuenta de la variabilidad en el conjunto de observación y permiten su descripción a través de funciones de probabilidad. Una vez estimada, la función de probabilidad se emplea en el contexto de la teoría de decisión bayesiana para seleccionar la clase correcta. Por otra parte, las funciones discriminativas mapean directamente la observación al valor de una etiqueta de clase. El clasificador más representativo del esquema generativo son los modelos de mezcla de gaussianas y el clasificador representativo de la metodología de funciones discriminativas son las máquina de vectores de soporte. Además, se emplean diversos esquemas de combinación de clasificadores con el objetivo de sumar las fortalezas de ambas metodologías de clasificación.

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. Para el problema de clasificación binario se utilizó una base de datos conformada por 483 señales Vulcano-Tectónicas (VT) y 580 señales Largo-Período (LP); en el caso de la clasificación multiclase, se realizaron las pruebas con una base de datos que está compuesta por 782 señales relacionadas con hielo (IC), 580 señales Largo- Período (LP) y 487 señales Vulcano-Tectónicas (VT).

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).

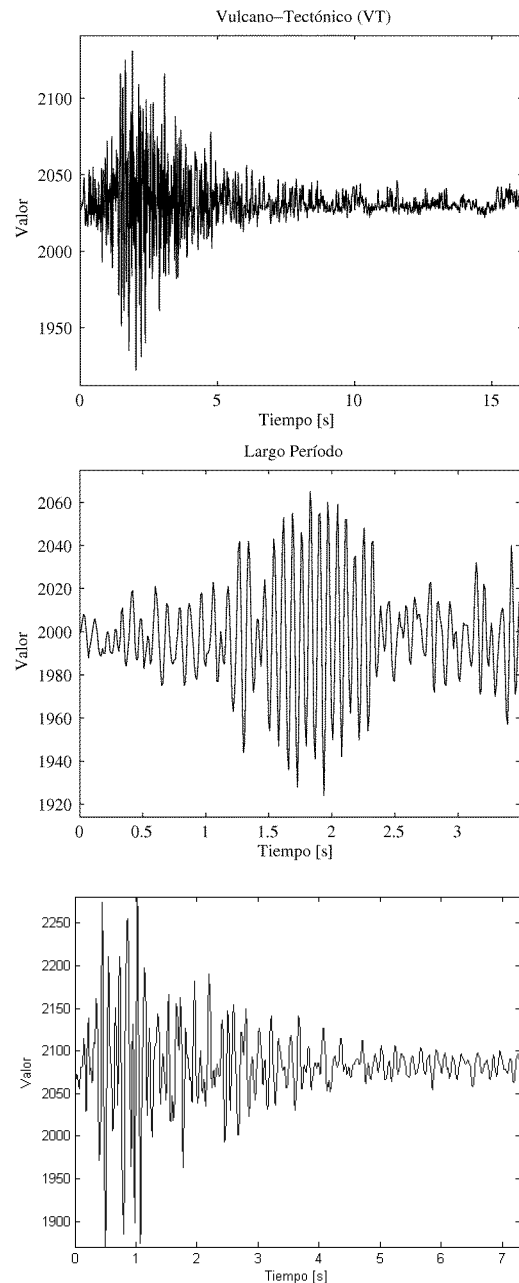


Figura 1. Formas de onda características de las señales de las clases VT, LP e IC respectivamente

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. ESQUEMAS DE CLASIFICACIÓN

La solución al problema de decisión se puede enfocar desde tres puntos de vista diferentes [4], [5]:

a. Para el primer caso, se resuelve el problema de inferencia determinando las densidades de clase condicionales $p(\mathbf{x}|C_k)$ para cada clase C_k , siendo $\mathbf{x} = (x_1, x_1, \dots, x_N)^T$ el vector que contiene las N observaciones, de forma individual. También de forma separada se infiere la probabilidad de clase a priori $p(C_k)$. Luego, a partir del teorema de Bayes

$$p(C_k | \mathbf{x}) = \frac{p(\mathbf{x} | C_k)p(C_k)}{p(\mathbf{x})}$$

se encuentra la probabilidad posterior de clase $p(C_k | \mathbf{x})$. En forma general, el denominador en el teorema de Bayes se especifica en términos de las cantidades que aparecen en el numerador, así:

$$p(\mathbf{x}) = \sum_{k=1}^K p(\mathbf{x} | C_k)p(C_k)$$

La probabilidad conjunta $p(\mathbf{x}, C_k)$ se puede modelar directamente y luego se normaliza para obtener las probabilidades posteriores. Teniendo las probabilidades posteriores, se utiliza la teoría de decisión para determinar la pertenencia a una clase de cada nueva entrada \mathbf{x} . Esta forma de solución se conoce como modelo generativo, ya que permite modelar explícita o implícitamente la distribución de las entradas como las salidas; y a partir de muestreo, se pueden generar datos sintéticos en el espacio de entrada.

b. En el segundo caso, el problema de inferencia se resuelve determinando las probabilidades posteriores de clase $p(C_k | \mathbf{x})$, y luego se utiliza la teoría de decisión para asignar cada nuevo \mathbf{x} a una de las clases. Este tipo de solución se conoce como modelos discriminativos, pues utilizan directamente la probabilidad posterior.

c. Finalmente, se puede encontrar una función $f(\mathbf{x})$, llamada función discriminante, que realice un mapeo de cada entrada \mathbf{x} directamente sobre una etiqueta de clase. En el caso de un problema de 2 clases, $f(\cdot)$ puede entregar un valor binario de forma que $f = 0$ represente la clase C_1 y $f = 1$ representa la clase C_2 . En este caso no se utiliza la teoría de probabilidad.

En el caso de la primera solución se debe considerar que el costo computacional es mayor dado que se debe encontrar la probabilidad conjunta sobre \mathbf{x} y C_k . En algunos casos \mathbf{x} presenta alta dimensionalidad y esto implica contar con un conjunto de entrenamiento grande. Las probabilidades de clase pueden ser estimadas de la fracción de los datos del sistema de entrenamiento que

conforman cada clase. Una gran ventaja de esta solución es que se puede utilizar la densidad marginal de los datos para detectar que nuevos datos tienen una baja probabilidad dado el modelo, estos datos hacen que la predicción pierda exactitud.

5. CLASIFICADORES

Modelo de mezclas de gaussianas

Los modelos de mezclas de Gaussianas (GMM Gaussian Mixture Models) han mostrado ser una herramienta poderosa para distinguir fuentes acústicas con diferentes propiedades generales. En reconocimiento de hablante, esta habilidad se ha explotado comúnmente, modelando cada hablante con un GMM [6]. Un GMM está compuesto, básicamente, de una superposición de K funciones de densidad de probabilidad (fdp) gaussianas, donde cada fdp está ponderada por un coeficiente de peso

$$p(\mathbf{x}) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x} | \mathbf{i}_k, \mathbf{O}_k)$$

donde \mathbf{i}_k y \mathbf{O}_k son la media y la matriz de covarianza de la función de densidad gaussiana asociada a la componente de mezcla k .

Por cada clase se estiman los parámetros de los GMM, que incluyen los coeficientes de ponderación, y las medias y matrices de covarianza de cada fdp gaussiana. Los parámetros del modelo suelen estimarse empleando el algoritmo EM, que es una versión generalizada de la estimación de máxima verosimilitud. Las ecuaciones de actualización están dadas por [5]:

$$\pi_k^{nuevo} = \frac{N_k}{N}$$

$$\mathbf{i}_k^{nuevo} = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) \mathbf{x}_n$$

$$\mathbf{O}_k^{nuevo} = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) (\mathbf{x}_n - \mathbf{i}_k^{nuevo})(\mathbf{x}_n - \mathbf{i}_k^{nuevo})^T$$

donde

$$N_k = \sum_{n=1}^N \gamma(z_{nk}), \quad \gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(\mathbf{x}_n | \mathbf{i}_k, \mathbf{O}_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(\mathbf{x}_n | \mathbf{i}_j, \mathbf{O}_j)}$$

En el paso E del algoritmo se calcula $\gamma(z_{nk})$ empleando un conjunto de parámetros iniciales y en el paso M, los parámetros se estiman usando las ecuaciones de actualización anteriores.

Máquina de vectores de soporte

La introducción de las máquinas de vectores de soporte (SVM) es uno de los desarrollos más recientes y promisorios en el área de aprendizaje de máquina [8]. Las máquinas de vectores de soporte fueron propuestas originalmente como un clasificador de dos clases: en donde, a partir de n puntos de entrenamiento etiquetados

$\{\mathbf{x}_i, y_i\}_{i=1}^n$; $y_i \in \{-1, 1\}$; se obtiene la SVM como el hiperplano $f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_0$ que maximiza la margen entre las clases $2/\|\mathbf{w}\|^2$, que se define, por la distancia de los patrones de entrenamiento más cercanos (a los que se les da el nombre de vectores de soporte).

El hiperplano descrito puede ser extendido a funciones de decisión no lineales a partir de un mapeo $\mathbf{Ö}$ a un espacio dimensional más grande, de esta forma, el clasificador lineal se construye como sigue [9]:

$$f(\mathbf{x}) = g(\mathbf{Ö}(\mathbf{x})) = \mathbf{w}^T \mathbf{Ö}(\mathbf{x}) + w_0$$

En la práctica, se reemplaza el mapeo $\mathbf{Ö}$ por un operador kernel K ; en consecuencia, la función de decisión puede ser escrita como sigue:

$$f(\mathbf{x}) = \sum_{\forall \mathbf{x}_i \in S} \alpha_i y_i K(\mathbf{x}, \mathbf{x}_i) + \alpha_0$$

Con S como el sistema de vectores de soporte, $y_i = \pm 1$ y $\alpha_i > 0$ son los parámetros que se deben optimizar en la etapa de entrenamiento.

6. RESULTADOS

Los experimentos fueron realizados para comparar el desempeño de los clasificadores con enfoques generativo y discriminativo entrenados con las representaciones espectrales mencionadas anteriormente (FFT, MCOV, PBURG y YULEAR). Se utilizó el criterio de validación cruzada con 5 particiones.

Clasificación binaria

En el caso del clasificador basado en un modelo de mezclas gaussianas, se eligió como matriz de covarianza una diagonal, debido al número de muestras por clase no se considera suficiente para trabajar con una matriz de covarianza densa. El criterio para elegir el número de componentes de mezcla fue el error de clasificación, obteniendo como valor óptimo un número de componentes igual a 2. En la figura 2, se puede apreciar como el aumento del número de componentes afecta negativamente la precisión para el caso particular de la representación espectral YULEAR.

Para el clasificador basado en máquinas de vectores de soporte, se utilizó como kernel, una función de base radial de la forma $\exp(-\gamma|u-v|^2)$, con factor de regularización igual a 1 y un valor de γ igual a 0.003 encontrado experimentalmente. En la figura 3 se muestra el comportamiento de la precisión contra el valor de γ dado en escala logarítmica.

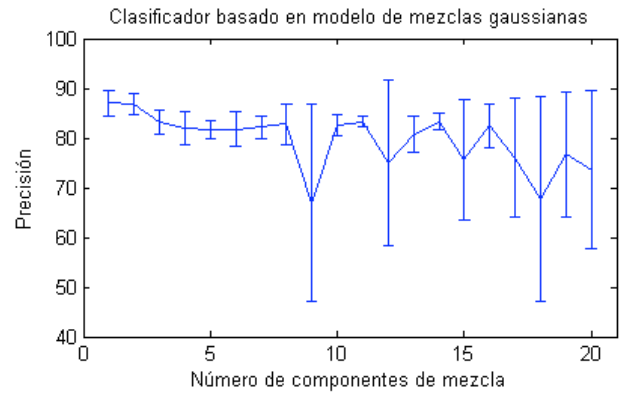


Figura 2. Relación entre precisión y número de componentes de mezcla para un clasificador GMM binario.

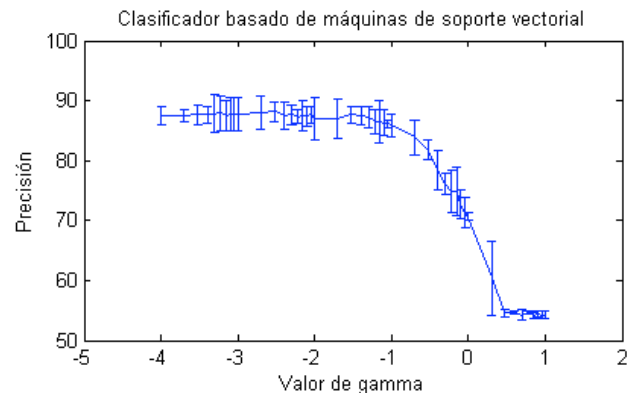


Figura 3. Relación entre precisión y el valor de γ para un clasificador SVM binario.

Los algoritmos utilizados fueron implementados para el caso de GMM utilizando Netlab [16] y para el caso de SVM, con LIBSVM [17]. Los resultados obtenidos para ambos clasificadores utilizando los parámetros mencionados con anterioridad se muestran en la tabla 1.

Representación Espectral	Clasificadores	
	GMM	SVM
FFT	80.38±2.53	83.21±3.71
MCOV	86.69±1.74	87.83±4.02
PBURG	86.41±1.64	87.92±2.68
YULEAR	86.79±2.02	88.21±1.66

Tabla 1. Resultados obtenidos con ambos clasificadores para las representaciones espectrales empleadas, FFT, MCOV, PBURG y PYULEAR.

En la tabla 1 de resultados, se puede observar que el clasificador SVM presenta un mejor desempeño en cuanto a precisión de clasificación con respecto al clasificador GMM. En particular, el mejor resultado en términos de precisión (y desviación estándar) se obtiene con la representación basada en estimación espectral con el método de Yule-Walker.

Clasificación multi-clase

En el caso de clasificador basado en un modelo de mezclas gaussianas, se eligió como matriz de covarianza una esférica. El número de componentes de mezcla se encontró de forma experimental, obteniendo, a partir de la precisión, que el valor de componentes de mezcla es igual a 1. En la figura 3 se muestra la variación de la precisión con respecto al número de componentes para el caso de la representación espectral PBURG.

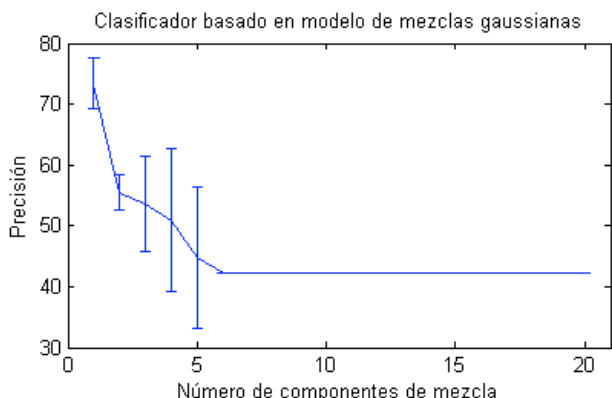


Figura 4. Relación entre precisión y número de componentes de mezcla para un clasificador GMM multi-clase.

Para el clasificador basado en máquinas de vectores de soporte, se obtuvieron los mejores resultados utilizando un kernel polinomial, de orden 3 y con factor de penalización igual a 1. En la figura 4 se muestra el comportamiento de la precisión contra el orden del polinomio para 3 valores diferentes.

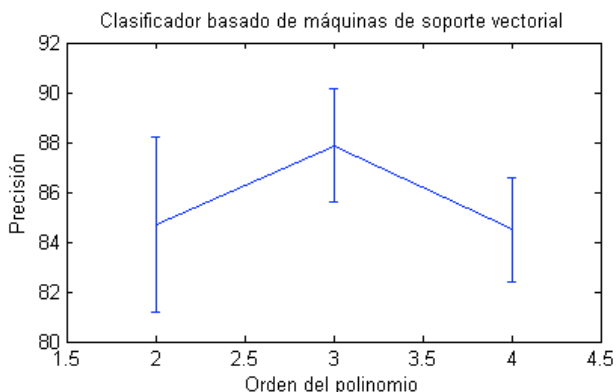


Figura 5. Relación entre precisión y el valor de γ para un clasificador SVM multi-clase.

En la tabla 2 se puede observar que de nuevo, el clasificador SVM presenta un mejor desempeño en cuanto a precisión de clasificación con respecto al clasificador GMM. En particular, el mejor resultado en términos de precisión (y desviación estándar) se obtiene con la representación basada en estimación espectral con el método de Burg.

Representación Espectral	Clasificadores	
	GMM	SVM
FFT	46.34±1.78	83.08±3.36
PBURG	73.33±4.18	87.86±2.26
YULEAR	70.00±3.74	87.64±1.93

Tabla 2. Resultados obtenidos con ambos clasificadores para las representaciones espectrales empleadas, FFT, PBURG y PYULEAR

Combinación de clasificadores

Adicional a los enfoques de clasificación propuestos, se implementó un esquema de clasificación a partir de estrategias de combinación de clasificadores basadas en reglas [18], con miras a mejorar la precisión resultante de clasificación. Luego de aplicar combinación de clasificadores empleando regla del máximo, mínimo y de mediana, se obtuvieron los resultados que se muestran en la tabla 4, para el caso binario, evidenciándose que ninguna de las reglas de combinación hace posible mejorar el desempeño individual de los clasificadores considerados. Sin embargo, se puede reducir ligeramente la desviación estándar, lo que sugiere desde el punto de vista estadístico, una mayor estabilidad del resultado.

Señal	Regla de combinación		
	Max	Min	Mediana
FFT	81.32±2.27	81.32±2.27	81.22±2.24
PMCOV	86.69±1.61	86.69±1.61	86.60±1.58
PBURG	87.07±1.43	87.07±1.43	86.98±1.55
YULEAR	86.69±2.52	86.69±2.52	86.69±2.52

Tabla 4. Resultados obtenidos con estrategias de clasificación para clasificación binaria

Repitiendo el procedimiento para el experimento de clasificación multi-clase, de la tabla 6 se puede observar que luego de la combinación, el clasificador resultante se degrada sustancialmente en su desempeño. Los resultados indican que las reglas de combinación en este caso en particular (multi-clase) van en detrimento de los resultados individuales.

Señal	Regla de combinación		
	Max	Min	Mediana
FFT	62.65±1.78	62.17±1.60	62.49±1.68
PBURG	30.95±3.65	30.84±3.74	30.54±3.56
YULEAR	63.74±2.68	63.95±2.75	63.19±2.60

Tabla 6. Resultados obtenidos con estrategias de clasificación para clasificación multi-clase

7. CONCLUSIONES

En este artículo se ha mostrado que para el caso de clasificación de eventos sísmicos empleando representaciones espectrales, las máquinas de vectores de

soporte proveen mejores resultados tanto en cuanto a precisión como a la desviación estándar producto de la validación cruzada, sobre todo para el caso multi-clase que requiere de una mayor estructura por parte de la extracción de características y de la clasificación para llevarse a cabo satisfactoriamente. De manera adicional, en el caso de dos clases, YULEAR fue la representación que condujo a mejores resultados y en el caso multi-clase de otro lado, la representación denominada PBURG. Respecto de la caracterización, no es una coincidencia que la mejor representación varíe según el caso, debido a que es posible que con la adición de la clase IC, el clasificador requiera de información no provista por YULEAR, o en su defecto, no resaltada lo suficiente para ser discriminativa desde el punto de vista de los clasificadores.

Con respecto al clasificador basado en mezclas de gaussianas, partiendo de los resultados obtenidos y del comportamiento del número de mezclas de las figuras 4 y 2, se tiene que, debido a que conforme el número de mezclas crece, la precisión decrece, por tanto, el número de muestras utilizadas para realizar el entrenamiento del modelo no es suficiente, resultando en una precisión de clasificación pobre. Una posible solución a este problema consistiría en aumentar substancialmente el conjunto de entrenamiento con miras a explotar la capacidad del modelo.

Es importante anotar que dado que las reglas de combinación de clasificadores no permiten mejorar las tasas de clasificación, y en el caso multi-clase las reducen considerablemente; lo más probable es que existan eventos mal etiquetados, ya que las reglas de combinación empleadas asumen que dado que los clasificadores son de diferente naturaleza, no deberían incurrir siempre en los mismos errores [18]. Una posible utilización del fenómeno observado en este trabajo, sería cambiando de perspectiva, la de desarrollar una metodología que permitiera el descubrimiento de eventos mal etiquetados a partir de los clasificadores, las reglas empleadas y posiblemente, en conjunto con técnicas basadas en agrupamiento y de descubrimiento de novedades en observaciones [5].

8. 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.

Los autores agradecen a Mauricio Orozco, estudiante de Doctorado, Universidad Nacional de Colombia, sede Manizales, por su colaboración en la realización de este artículo.

9. 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. Bishop. Pattern Recognition and Machine Learning. Springer. 2006.

[6] FALTHAUSER R., PFAU, T. y RUSKE G., "On-line speaking rate estimation using Gaussian mixture models". *Proceedings of the International Conference in Acoustics, Speech and Signal Processing*, 2000, pp. 1355–1358.

[7] M. Tipping, C. Bishop. Mixtures of principal component analyzers. Neural Computation. 11(2). 1999. J. Shawe-Taylor, N. Cristianini. Kernel Methods for Pattern Analysis. Cambridge University Press. 2004.

[8] V. Vapnik. Statistical Learning Theory. John Wiley & Sons. 1998.

[9] M. Orozco-Alzate, M.E García-Ocampo, R.P.W. Duin, C.G. Castellanos. Dissimilarity-Based Classification of Seismic Volcanic Signals at Nevado del Ruiz Volcano. Earth Sciences Research Journal, vol. 10, No. 2, pages 57-65, 2006.

[10] T. Fawcett. An Introduction to ROC Analysis. Pattern Recognition Letters. Vol. 27. 2006. pp. 861-874.

[11] M. Orozco-Alzate, M.E García-Ocampo, R.P.W. Duin, C.G. Castellanos. Dissimilarity-Based Classification of Seismic Volcanic Signals at Nevado del Ruiz Volcano. 2nd Latin American Congress of Seismology, 2006.

[12] 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.

[13] B. Schölkopf, A. Smola. Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond. The MIT Press. 2001.

[14] D. MacKay. Information Theory, Inference & Learning Algorithms. Cambridge University Press. 2002.

[15] C.E. Rasmussen y C.K.I. Williams. Gaussian Processes for Machine Learning. The MIT Press. 2006.

[16] Netlab: toolbox de reconocimiento de patrones para Matlab. Disponible en: <http://www.prtools.org>.

[17] Ch.Chang and Ch. Lin, LIBSVM : librería para máquinas de vectores de soporte, 2001. Software disponible en: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.

[18] J. Kittler, M. Hatef, R. Duin, and J. Matas. On Combining Classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998.