

Identificación de estados funcionales en bioseñales empleando Máquinas de Soporte Vectorial

Ricardo Henao¹, Fabián Ojeda¹, Mauricio Orozco¹, and Germán Castellanos¹

Universidad Nacional de Colombia, G. Control y Procesamiento Digital de Señales,
Campus La Nubia Manizales, Colombia

{rhenao, fabianojeda, mauricio_orozco, gcastell}@unalmz1.edu.co
<http://www.manizales.unal.edu.co>

Resumen Se presenta el empleo de máquinas de soporte vectorial, las cuales están basadas en el principio de minimización de riesgo estructural y que garantizan alta capacidad de generalización. Las SVM encuentran amplio uso en las tareas de reconocimiento. Sin embargo, la adecuación de la estructura de la SVM en cada caso concreto puede ser una de las limitantes en su uso. Se propone una metodología de desarrollo de las SVM aplicada a la identificación de patologías en señales ECG y voz. En este trabajo se estudiará la identificación de los estados funcionales de normalidad y anormalidad presentes en los conjuntos de rasgos dominantes de los registros de voz y ECG, por medio de las SVM. Se presentan e ambos caso resultados de comparación con otras técnicas de reconocimiento.

1 Introducción

Actualmente, se realizan numerosos estudios orientados al análisis de la efectividad en la aplicación de las técnicas clásicas estadísticas, así como de los métodos contemporáneos basados en inteligencia artificial, para la clasificación de bioseñales (voz, ECG, EEG, etc.) en clases de diversos grupos correspondientes a estados funcionales del cuerpo: normalidad y patologías. Entre las técnicas disponibles para la identificación de los estados funcionales corporales se encuentran las siguientes:

- *Técnicas clásicas de discriminación.* Basadas en el índice de discriminación de Fisher, estas técnicas tiene el inconveniente de ser aplicables a conjuntos de rasgos linealmente separables.
- *Técnicas Bayesianas.* En este caso se supera la limitación anterior. Sin embargo, un defecto de este enfoque reside en la necesidad de postular una densidad de probabilidad multivariante de los rasgos, para lo cual la información usual no es suficiente.
- *Redes neuronales.* Algunas redes neuronales, tales como el preceptor multi-capas y las funciones de base radial, son altamente idóneas para propósitos de clasificación. Tienen además la ventaja de acoplarse bien con las técnicas

Bayesianas, por lo que a sus estimaciones se les puede dar una interpretación probabilista.

- *Máquinas de soporte vectorial*. Sobre el método Bayesiano, esta técnica presenta la ventaja de no requerir ningún tipo de hipótesis sobre la densidad de probabilidad de los rasgos, mientras que sobre las redes neuronales ofrecen la prerrogativa de ser convenientes en términos de la dimensionalidad del problema.

Las SVM han sido aplicadas con éxito en problemas tan diversos como el análisis de imágenes, la estimación de densidades de probabilidad y el estudio de confiabilidad de estructuras, en general, estas resultaron ser el cuarto método más utilizado para clasificación después de MLP, CART y KNN. Sin embargo, es frecuente el *mal empleo* que se les da probablemente por falta de conocimiento de la metodología de su diseño [5].

En este trabajo se estudiará la identificación de los estados funcionales de normalidad y anormalidad presentes en los conjuntos de rasgos dominantes de los registros de voz y ECG, por medio de las SVM. Se presentan en ambos casos resultados de comparación con otras técnicas de reconocimiento.

2 Máquinas de Soporte Vectorial. Fundamentos

Las máquinas de soporte vectorial (SVM) se emplean en el reconocimiento de patrones, particularmente, en tareas de clasificación. Sustentadas en el principio de minimización de riesgo estructural (SRM), éstas garantizan buena capacidad de generalización, debido a que el subconjunto de funciones encontradas en el proceso de optimización minimiza el riesgo actual del problema, de manera que entrenando una serie de máquinas para un objetivo dado se minimiza el riesgo empírico y la confianza de la dimensión VC [10].

La cantidad $R(\alpha)$, llamada riesgo no posee formulación exacta, pero puede ser expresada como una cota. El riesgo empírico $R_{emp}(\alpha)$ definido solamente como el promedio de los errores de entrenamiento para un número finito y fijo de observaciones $\{x_i, y_i\}$.

$$R_{emp}(\alpha) = \frac{1}{2l} \sum |y_i - f(x_i, \alpha)|$$

La cantidad o *pérdida* $\frac{1}{2} |y_i - f(x_i, \alpha)| \in [0, 1]$. Para un número $\eta \in (0, 1)$, que representa las pérdidas, se mantiene que [10]:

$$R(\alpha) \leq R_{emp}(\alpha) + \sqrt{\left(\frac{h(\log(2l/h) + 1) - \log(\eta/4)}{l}\right)}$$

donde h es una constante no negativa llamada dimensión Vapnik-Chervonenkis (VC) que es la medida de la capacidad o *confidencia* VC.

Sea dado un grupo de datos de entrenamiento $\{x_i, y_i\}$ con $i = 1, \dots, l$, $y_i \in \{-1, 1\}$ y $x_i \in \mathbb{R}^d$. Existe un hiperplano que separa los datos de etiquetas positivas y negativas, tales que:

$$\begin{aligned}
x_i \cdot w + b &\geq 1 - \xi_i \text{ para } y_i = 1 \\
x_i \cdot w + b &\leq -1 + \xi_i \text{ para } y_i = -1 \\
\xi_i &\geq \forall i
\end{aligned}$$

donde w es la normal al hiperplano y ξ_i son las variables introducidas por errores de clasificación en calidad de violaciones del hiperplano, de manera que $\sum \xi_i$ sea la cota del error de clasificación. Una manera directa de añadir el costo a la función objetivo es minimizar $\|w\|^2/2 + C \sum \xi_i$, siendo C la constante elegida, correspondiente al inverso del valor de la penalización de los errores. Así, se tiene un caso de optimización convexa como problema de programación cuadrática (QP), cuya forma dual Wolfe está dada por [9]:

Maximizar:

$$L_D \equiv \sum \alpha_i - \frac{1}{2} \sum \alpha_i \alpha_j y_i y_j x_i x_j$$

Sujeto a:

$$0 < \alpha_i < C$$

$$\sum \alpha_i y_i = 0$$

con solución:

$$w = \sum^{N_s} \alpha_i y_i x_i$$

donde N_s es el número de vectores de soporte.

Por cuanto en la mayoría de los casos, el espacio de entrada no es lineal, es necesario hacer su transformación basados en el producto punto, en orden a mapearlos a un espacio euclídeo \mathcal{H} de manera que:

$$\Phi : \mathbb{R}^n \rightarrow \mathcal{H}$$

Por consiguiente, el algoritmo de entrenamiento solo depende de los datos de entrada a través de los productos punto de la forma $\Phi(x_i) \cdot \Phi(x_j)$. Existirá, entonces, una función K llamada *kernel*, tal que se cumpla $K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$. Tal vez, el kernel más utilizado es el RBF (*Radial Basis Function*) definido como [1]:

$$k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$$

que en particular es infinito dimensional en \mathcal{H} y satisface las condiciones de Mercer. Otros kernels propuestos son los siguientes [9]:

$$k(x_i, x_j) = (\gamma x_i \cdot x_j + 1)^p$$

$$k(x_i, x_j) = \tanh(-\gamma(x_i \cdot y_j) - \delta)$$

donde el primero es llamado *polinomial* y el segundo MLP (*Multilayer Perceptron*), debido a que es una particularización de la red neuronal sigmoideal de dos capas.

3 SVM en tareas de clasificación

En el diseño de las SVM es importante tener en cuenta los siguientes elementos:

3.1 Capacidad de generalización

Dado que las SVM no minimizan una superficie de error, sino un margen que mide la separación entre las clases, el problema de optimización es convexo, por lo tanto siempre tiene un mínimo global. Para esto, las SVM se orientan en la obtención de dicho mínimo mediante el algoritmo de la función dual de Wolfe [9]. Así mismo, en casos en los que se asegure la separación máxima entre las clases, las máquinas sugieren el hiperplano óptimo que garantice buena capacidad de generalización en un esquema de clasificación correctamente establecido, esto es, un espacio de entrada sólido con características relativamente invariantes.

3.2 Funciones Kernel y espacios no lineales

El uso del kernel en las SVM corresponde a una función de decisión que no sea linealmente dependiente del espacio de entrada. Convenientemente, al reemplazar la función kernel en la SVM genera una máquina que exista en un espacio infinito dimensional y, que se toma la misma cantidad de tiempo en la fase de entrenamiento que otras con datos no mapeados. La única restricción de este procedimiento, consiste en que el kernel debe cumplir las condiciones de Mercer, en el sentido en que deben llenar todo el espacio euclídeo de Hilbert, es decir, ser positivos definidos. En este sentido, para ciertos valores, el kernel MLP no es positivo definido.

3.3 Arquitectura

La arquitectura de las SVM, solo depende del parámetro C , la función kernel (incluyendo sus parámetros), que para el caso del RBF [1] solo se requiere el parámetro σ , evitando así, requerimientos sobre parámetros exclusivos de arquitectura, tales como número de nodos y capas, tipo de conexión entre capas, etc. Aunque, en la literatura se pueden encontrar métodos heurísticos para la selección de dichos parámetros, o bien, métodos automáticos con restricciones para el tipo de kernel, incluyendo el número de clases y de parámetros a seleccionar.

Sobre la selección del modelo, variados métodos han sido planteados en este caso. Así, en [4] se propone un procedimiento heurístico de selección para el kernel tipo RBF, basado en los resultados de una SVM lineal,

3.4 Algoritmos de optimización y costo computacional

Entre los algoritmos de optimización conocidos, el SMO [8] es el más utilizado debido a su efectividad computacional, el cual descompone el problema de optimización en tareas mucho más pequeñas, en orden a reducir el tamaño de las operaciones matriciales, es decir, requerimientos de memoria y procesador. En [3] se analiza la convergencia de dicho algoritmo y en [2] se plantean modificaciones para mejorar su rendimiento. En cualquier caso, los algoritmos de optimización en las SVM, mejoran su rendimiento y son superiores en problemas con grandes espacios de entrada, donde muchas metodológicas fallan por convergencia o costo computacional, por ejemplo, las RNA y los CART.

3.5 Validación del error

En cuanto a la validación del error, las dos formas más utilizadas son: *a)* la prueba simple independiente (IDT *independent test sample*), que consiste en partir el espacio de entrada en dos, uno para la fase de entrenamiento y otro para la de validación con el fin de medir el desempeño del clasificador en cuanto a generalización y, *b)* La validación cruzada (CV *cross validation*), cuando varias particiones sirven a la vez de entrenamiento y validación, pero de manera alternada. La primera, es bastante común en problemas con grupos de datos relativamente grandes, en cambio la segunda, en general para grupos pequeños, cuando se requiere una validación más estricta con la desventaja de incrementar el costo computacional, debido a que el número de fases de entrenamiento se incrementa tanto como la cantidad de particiones de los datos. De este modo, en cualquiera de los dos casos, son requeridos una buena capacidad de generalización y un algoritmo eficiente de entrenamiento.

4 Identificación de estados funcionales en bioseñales

El área del reconocimiento de señales biomédicas y, más específicamente, la parte de clasificación de las mismas ha sido durante los últimos años un tema de investigación muy atractivo debido a los avances de las técnicas de reconocimiento de patrones y la posibilidad de generar diagnósticos y análisis sin la utilización de técnicas invasivas que pueden ser inútiles, costosas, inoficiosas y peligrosas para pacientes o individuos en observación. Aunque la teoría de clasificadores se ha ido depurando sobre todo en los últimos 5 años, todavía es común encontrar material de investigación con técnicas hoy por hoy caídas en desuso como: clasificadores bayesianos, *k*-ésimo vecino cercano simple (KNN), mapas de teoría de resonancia adaptativa sencillos (ART1) y perceptrones multicapa (MLP) con algoritmos de entrenamiento de gradiente descendiente conjugado sin inicialización de pesos.

4.1 Clasificación de señales ECG

Se utilizó un método para clasificar señales de ECG con extracción de características basada en la información contenida en los coeficientes de aproximación

de su descomposición wavelet. El estudio incluyó 11 formas de onda etiquetadas de la base de datos de arritmias MIT-BIH: Latido auricular prematuro desviado (a), latido de escape ventricular (E), fusión de latido ventricular y normal (F), fusión de latido acelerado y normal (f), bloqueo de rama izquierda (L), latido normal (N), latido acelerado (P), onda P no conducida (p), bloqueo rama derecha (R), contracción ventricular prematura (V) y fibrilación ventricular (VF).

Tabla 1. Resultados para validación cruzada de 5 particiones

Conjunto	C	γ	Precisión (%)	Desviación (%)
Sin PCA	8	0.4	97.2727	11.82188
PCA 95 %	8	0.5	97.0909	11.85367
PCA 90 %	8	0.8	96.7273	11.82541

En la tabla 1 se presentan los resultados correspondientes a la clasificación usando SVM para un conjunto de características obtenido con una descomposición de cuarto nivel utilizando la wavelet Daubechies 2 (db2). Todas las pruebas restantes para niveles de descomposición 3 y 4, Wavelets ortogonales (daubechies-dbN, symlets-symN y coiflets-coifN), pares de Wavelets biortogonales (biorNr.Nd y rbioNr.Nd), selección efectiva de características usando PCA y reconocimiento con clasificadores bayesianos y redes neuronales artificiales puede encontrarse en [7]. Sin embargo, en el mejor de los casos para Bayes y redes neuronales los resultados no superaron el 94.55 %.

4.2 Clasificación de señales de voz

Las muestras de voz utilizadas, fueron obtenidas de una población adulta de la ciudad de Manizales, donde se evaluaron pacientes de ambos sexos con edades entre los 18 y 62 años. El número total de registros analizados fue de 91, todos valorados por un fonoaudiólogo. Cada muestra consiste en un registro de la fonación sostenida de los cinco fonemas vocálicos. Estas señales fueron grabadas con un micrófono dinámico, colocado a 4cm de los labios del paciente, digitalizados a 22050Hz y 16 bits por muestra. Las clases de voz a identificar son: *voz normal* y *voz patológica*. Se conformaron dos conjuntos. El primero de ellos usando el fonema /a/, mientras el segundo, conformado por todas la vocales. Las características calculadas, son representadas por doce coeficientes wavelet obtenidos de cada fonema.

Tabla 2. Resultados para validación cruzada de 5 particiones

Conjunto	C	γ	Precisión (%)	Desviación (%)
A	4	0.04	93.4	8.6317
B	4	0.06	99.9	0.00

Tabla 3. Resultados SVM utilizando PCA

Conjunto	Varianza acumulada	Reducción	SVM			
			C	γ	Accuracy(%)	Std(%)
A	90%	3/12	4	0.04	85.7143	4.5185
	95%	4/12	4	0.04	93.4066	8.632
	99%	6/12	4	0.04	93.4066	8.632
B	87%	7/60	4	0.06	96.7033	3.042
	90%	9/60	4	0.06	99.99	0.00
	95%	12/60	4	0.06	99.99	0.00

Tomando en cuenta la naturaleza del problema, es decir, la consecución de un reducido conjunto de muestras (91 observaciones), dada la dificultad de obtener registros patológicos, es necesario el uso de una técnica de clasificación de alto desempeño en términos de su capacidad de generalización. Se utilizaron las siguientes: *perceptrones multicapa*, *fuzzy ARTMAP* y las *máquinas de soporte vectorial* (SVM). En las tablas (2) y (3), se presentan los resultados de clasificación para las SVM, los resultados para las restantes dos técnicas pueden encontrarse en [6]. Hay que anotar, que para este caso las SVM superaron a las otras técnicas utilizadas debido a que con ellas no se supero la cifra del 96.04 %.

5 Conclusiones

Se muestra la utilización de SVM en la clasificación de señales biomédicas, específicamente para ECG y Voz. Los resultados obtenidos son particularmente concluyentes sobre las fortalezas de las SVM: la capacidad de generalización y el comportamiento ante reducción de dimensionalidad.

En general, las pruebas muestran que en el caso de identificación de estados funcionales del cuerpo, cuando se tienen condiciones ambiguas en la extracción de características, la SVM es una buena elección en el sentido de que pueden aprovechar de mejor manera el espacio de características.

Sin embargo, el uso de un kernel, en este caso el gaussiano RBF sugiere una mejora en los resultados de clasificación y sujeta a una buena elección del modelo en el sentido de los parámetros del kernel utilizado, resultando en la solidez de la capacidad de generalización del clasificador. Basados en el mapeo infinito dimensional de los datos mediante el uso de kernel, es claro la necesidad de una menor cantidad de datos para la fase de entrenamiento sin sacrificar desempeño. El hecho de necesitar menos volumen de información para el entrenamiento reduce los requerimientos de proceso y memoria lo cual representa un ahorro sustancial de tiempo. En otros términos, la cantidad de datos que conforman un espacio de características es muy importante en el área de bioseñales debido a que la recolección de dichas señales es un proceso largo y sumamente costoso, por lo tanto, la posibilidad de obtener un clasificador confiable con un grupo reducido es primordial. Ejemplo de esto, es la aplicación mostrada de Voz en la cual el grupo de datos es particularmente reducido, sin embargo, los resultados muestran un alto rendimiento utilizando SVM.

Referencias

1. C.J.C. Burges. A tutorial on support vector machines for pattern recognition. *Knowledge Discovery and Data Mining*, 2:22, 1998.
2. S. Keerthi, S. Shevade, C. Bhattacharyya, and K. Murthy. Improvements to platt's smo algorithm for svm classifier design, 1999.
3. S. S. Keerthi and E. G. Gilbert. Convergence of a generalized SMO algorithm for SVM classifier design. *Machine Learning*, 46(1-3):351–360, 2002.
4. S.S. Keerthi and C.-J. Lin. Asymptotic behaviors of support vector machines with gaussian kernel. Technical report, Department of Computer Science, National Taiwan University, 2002.
5. Chih-Jen Lin. Can support vector machine be a major classification method ?, 2003.
6. F. Ojeda. Extracción de características usando transformada wavelet en la identificación de voces patológicas. Trabajo de grado, Universidad Nacional de Colombia, Manizales, 2003.
7. M. Orozco. Clasificación de arritmias cardíacas usando transformada wavelet y técnicas de reconocimiento de patrones. Trabajo de grado, Universidad Nacional de Colombia, Manizales, 2003.
8. J. Platt. Sequential minimal optimization: A fast algorithm for training support vector machines, 1998.
9. Bernhard Schölkopf and Alex Smola. *Learning with Kernels Support Vector Machines, Regularization, Optimization and Beyond*. MIT Press, Cambridge, MA, 2002.
10. V. Vapnik. *The Nature of Statistical Learning Theory*. Springer, N.Y., 1995.