Selección de características usando modelo híbrido basado en algoritmos genéticos

Feature selection using a genetic algorithm-based hybrid approach

Luis Felipe Giraldo,[1] Edilson Delgado Trejos,[2] Juan Carlos Riaño[3] y Germán Castellanos Domínguez[4]

RESUMEN

En el artículo se propone un modelo híbrido de selección de características con el objeto de reducir la dimensión del espacio de entrenamiento, sin comprometer la precisión de clasificación. El modelo incluye la inducción de un árbol de decisión que genera subconjuntos de características, para las cuales seguidamente se evalúa su relevancia mediante el criterio del mínimo error de clasificación. El procedimiento de evaluación se desarrolla empleando la regla de los *k*-vecinos más cercanos. Usualmente, la reducción de espacios supone una cota de error de clasificación; sin embargo, en este trabajo la sintonización del modelo híbrido de selección se realiza usando algoritmos genéticos, con lo cual se obtiene de forma simultánea la minimización tanto del número de características de entrenamiento, como del error de clasificación. De manera adicional, a diferencia de las técnicas convencionales de selección, el modelo propuesto permite cuantificar el nivel de relevancia de cada característica perteneciente al conjunto reducido de entrenamiento. Las pruebas del modelo se realizan para la identificación de hipernasalidad, en el caso de voz, y cardiopatía isquémica, en el caso de registros de electrocardiografía. Las bases de datos corresponden a una población de 90 niños (45 registros por clase) y a 100 registros electrocardiográficos (50 por clase). Los resultados obtenidos muestran una efectividad promedio para la reducción del espacio de entrenamiento inicial hasta de un 88%, con una tasa promedio de error de clasificación inferior al 6%.

Palabras clave: selección de características, algoritmos genéticos, árboles de decisión, *k*-vecinos más cercanos, relevancia.

ABSTRACT

The present work proposes a hybrid feature selection model aimed at reducing training time whilst maintaining classification accuracy. The model includes adjusting a decision tree for producing feature subsets. Such subsets' statistical relevance was evaluated from their resulting classification error. Evaluation involved using the k-nearest neighbors' rule. Dimension reduction techniques usually assume an element of error; however, the hybrid selection model was tuned by means of genetic algorithms in this work. They simultaneously minimise the number of features and training error. Contrasting with conventional methods, this model also led to quantifying the relevance of each training set's features. The model was tested on speech signals (hypernasality classification) and ECG identification (ischemic cardiopathy). In the case of speech signals, the database consisted of 90 children (45 recordings per sample); the ECG database had 100 electrocardiograph records (50 recordings per sample). Results showed average reduction rates of up to 88%, classification error being less than 6%.

Keywords: feature selection, genetic algorithm, decision tree, the k nearest neighbor rule, relevancy.

Recibido: agosto 1 de 2006
Aceptado: octubre 31 de 2006

Introducción

El reconocimiento de patrones empleando métodos computacionales requiere la elección de un espacio adecuado de entrenamiento que permita de manera suficiente la discriminación entre las clases. Sin embargo, en la mayoría de los casos, algunas de las características elegidas para formar dicha representación son irrelevantes. El procedimiento de

[1] Ingeniero electrónico, Universidad Nacional de Colombia. Candidato a M.Sc. en Ingeniería Electrónica, Universidad de los Andes. luispipe16@yahoo.com

[2] Ingeniero electrónico, Universidad Nacional de Colombia. M.Sc. en Ingeniería y Automatización Industrial, Universidad Nacional de Colombia. Candidato a Ph.D. en Ingeniería (L. I. Automática), Universidad Nacional de Colombia. edelgadot@unal.edu.co.

[3] Matemático, Universidad Nacional de Colombia. Candidato a Ph.D. en Automática, Universidad Nacional de Colombia. Profesor, Universidad Nacional de Colombia. jcrianoro@unal.edu.co.

[4] Ph.D. en Ingeniería, Nauchno Isseledovatelskiy Institut, MTUCI, Russia. Profesor titular, Universidad Nacional de Colombia, Manizales. gcastell@telesat.com.co.

selección de características consiste en la eliminación de aquellas con baja o sin ninguna relevancia, de tal manera que se reduzca el tiempo de aprendizaje del clasificador, mientras se mantiene un valor de precisión aceptable de clasificación, incrementando así la capacidad de generalización (Yu y Liu, 2004).

En la selección efectiva de características orientada al reconocimiento de patrones, cada vez es más frecuente el empleo de los algoritmos genéticos. Así por ejemplo, en (Hong *et al.,* 2006) se usan en conjunto con las redes neuronales para el reconocimiento de pacientes con cáncer, mientras en (Kim *et al*., 2004) se analiza su empleo en conjunto con árboles de decisión para el reconocimiento de defectos en imágenes. De otra parte, el empleo de los *k*-vecinos más cercanos como regla de clasificación permite la simplificación del procedimiento de entrenamiento (Peña, 2002). En particular, en (Raymer *et al*., 2000) se realiza la reducción de dimensión para pruebas bioquímicas y médicas, transformando geométricamente el espacio de características por medio del uso de algoritmos genéticos y *k*-vecinos más cercanos, con una reducción de características del 85%. Sin embargo, en los trabajos donde se proponen estos modelos híbridos de algoritmos genéticos con árboles de decisión, aunque se obtiene una alta reducción del número de características, no hay posibilidad alguna de disminuir significativamente el error de clasificación. En este sentido, se propone el empleo de un conjunto de pesos que modifiquen el espacio de características en la regla de clasificación, cuyos valores se sintonizan mediante el uso de algoritmos genéticos.

En el presente trabajo se desarrolla una técnica de selección de características que paralelamente minimiza el error de clasificación, empleando una métrica de representación que además proporciona un valor de relevancia para cada una de las características seleccionadas. En la segunda sección se describe la formulación teórica de la selección de características, así como la generación de subconjuntos mediante árboles de decisión, y la función de evaluación mediante la regla de los *k*-vecinos más cercanos. En la tercera sección se describen las bases de datos junto a la estructura metodológica para la selección de características. En la cuarta sección se exponen los resultados experimentales obtenidos, los cuales se comparan con otros modelos de selección de características propuestos en la literatura. Finalmente, en la quinta sección se presentan las conclusiones y el trabajo futuro.

Selección efectiva de características

De forma general, la selección efectiva de características contempla las siguientes tres etapas básicas de proceso [cerma]:

1. *Estrategias de generación*: se origina cada nuevo subgrupo de variables que va a ser analizado, tomando las características directamente del espacio inicial de entrenamiento.
2. *Función de evaluación*: mide la efectividad de cada subconjunto generado respecto a alguna métrica asociada a un criterio de relevancia. La función de evaluación incluye la *condición de parada*, que corresponde a la restricción impuesta sobre los valores umbrales de efectividad, cuya aparición implica la detención en la búsqueda de un siguiente posible grupo subóptimo de características.
3. *Validación*: tiene como objeto aceptar o descartar la consistencia de los subconjuntos encontrados respecto a la capacidad de representación orientada al reconocimiento de patrones.

Básicamente, referente a la reducción de dimensionalidad, existen los siguientes tipos de métricas (Bast, 2004):

- *Métricas geométricas*, orientadas a hallar los subespacios geométricos, con dimensión reducida frente al original, donde las variables rechazadas son las que no ofrecen discriminación entre clases en el espacio geométrico de representación.
- *Métricas estadísticas*, orientadas a generar modelos de representación de estructuras estadísticas que permitan describir con una menor cantidad de variables de representación el fenómeno aleatorio en análisis.
- *Métricas de información*, generan modelos de representación de la carga informativa de las variables, con el fin de construir estructuras relacionadas con la incertidumbre para descartar variables redundantes que sólo incrementan la complejidad y costos de cómputo.

Usualmente, además de asociarse a un criterio de relevancia para medir su efectividad en las tareas de clasificación, es necesario para las características hacer claro las posibles interacciones no lineales entre las variables de representación.

Relevancia

Definición 1 (*Relevancia*). Sean $\mathbf{k}=\{k_r:r=1,\ldots,L\}$ el conjunto de etiquetas de clase y $\xi=\{\xi_i:i=1,\ldots,p\}$ el conjunto de variables de representación, a partir del cual al extraer una características ξ_i se forma el subconjunto $\hat{\xi}_i=\xi-\xi_i$. Entonces, una variable de representación ξ_i es *relevante* respecto a la función dada de evaluación f_ξ, si y sólo si:

$$f(\mathbf{k},\xi)\neq f_{\hat{\xi}_i}\left(k,\hat{\xi}_i\right) \qquad (1)$$

Definición 2 (*Función de evaluación*). Sea $\hat{\xi}=\{\xi_i:\mathrm{i}=1,\ldots,\mathrm{M}\}$ un conjunto de características, donde se genera el espacio de entrenamiento $\mathbf{w}=\{w_i\in\mathbf{R}^M:i=1,\ldots,n\}$ a partir de los valores estimados obtenidos sobre las n observaciones. Sea $\mathbf{k}=\{k_r:r\in\mathbf{N}\}$ el conjunto de clases, de tal forma que por cada observación w_i sólo una clase, o etiqueta, es asignada. Se denomina *función de evaluación* f_ξ a aquella función, que de acuerdo con alguna métrica asociada, provee un valor correspondiente a los datos evaluados. Tal función puede ser descrita como:

$f_\xi:\mathbf{N}\times\mathbf{R}^M\rightarrow\mathbf{R}$, de forma que $(\mathbf{k},\mathbf{w})\rightarrow f_\xi(\mathbf{k},\mathbf{w})$.

Generación de subconjuntos por árboles de decisión

Un *árbol de decisión* consiste de una estructura jerárquica cuyos nodos representan las características; las ramas son los posibles valores para las características en cuestión y las hojas la clasificación del ejemplar. Un árbol de decisión clasifica un ejemplo, filtrándolo de manera descendente hasta encontrar una hoja, que corresponde a la clasificación buscada. En general, un árbol de decisión representa una disyunción de conjunciones que determinan las restricciones relacionadas a los posibles valores que pueden tomar las características de entrenamiento. Cada rama que va desde la raíz del árbol a una hoja representa una conjunción de tales restricciones, mientras el árbol mismo representa la disyunción de esas conjunciones.

Un procedimiento que desarrolla la búsqueda descendente y egoísta en el espacio de posibles árboles de decisión, corresponde al algoritmo ID3 (Quinlan, 1986), el cual evalúa cada una de las características con base en una función dada de evaluación, midiendo el rendimiento del clasificador durante el entrenamiento. Dicha característica es ubicada como nodo del árbol, donde se fragmentan los ejemplos para cada uno de los valores que puede tomar, siendo estos las ramas del árbol. El procedimiento se hace en forma recursiva hasta que los ejemplos de entrenamiento compartan la misma clase, ubicándose como una hoja del árbol. Esta rutina se ilustra en la Figura 1.

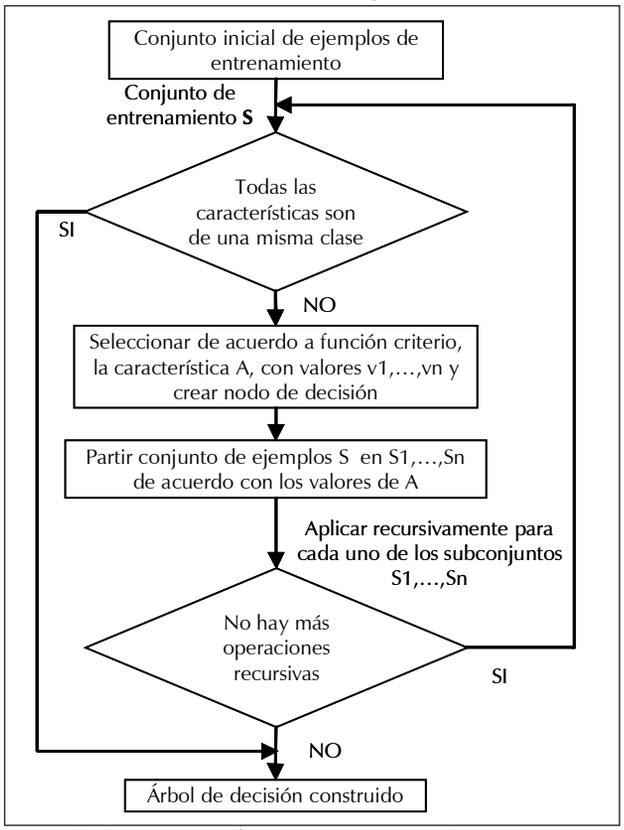


Figura 1. Diagrama de flujo del algoritmo ID3

La función de evaluación utilizada en el algoritmo ID3, propuesta para evaluar cada característica, es la ganancia de información o medida de la cantidad de información presente. En particular, se propone la entropía como una manera de cuantificar la bondad de cada característica en análisis. Así, sean los n posibles valores medidos para una característica $A=\{v_i:i=1,\ldots,n\}$, con probabilidad $P(v_i)$ de ocurrencia entonces, la entropía E de la respuesta actual se define como:

$$E=-\sum_{i=1}^{n}P(v_i)\log_2 P(v_i), \qquad (2)$$

Por lo tanto, la ganancia de información, que corresponde a la reducción de la entropía causada por fragmentar un conjunto de entrenamiento **S** respecto a una característica A, se define como:

$$\gamma(\mathbf{S},A)=E(\mathbf{S})-\sum_{v\in A}\frac{N(\mathbf{S}_v)}{N(\mathbf{S})}E(\mathbf{S}_v), \qquad (3)$$

donde $N(\mathbf{S})$ es el conjunto de ejemplos de entrenamiento para los cuales la característica A toma el valor v, mientras, $N(\mathbf{S}_v)$ es el tamaño del conjunto $\mathbf{S}_v$.

Sintonización de funciones de costo o de evaluación mediante algoritmos genéticos

Dado un método para codificar las posibles soluciones que puedan ser obtenidas para un problema dado, mediante n cromosomas de longitud L, y dada, además, una función de evaluación que proporcione una medida de evaluación, λ, para cada uno de los cromosomas, entonces el algoritmo genético se puede construir siguiendo el procedimiento descrito en el diagrama de flujo mostrado en la Figura 2 (Hong *et al.*, 2006). La función de evaluación que se propone usar es la regla de los k-vecinos más cercanos, debido a su capacidad de generalización (Peña, 2002).

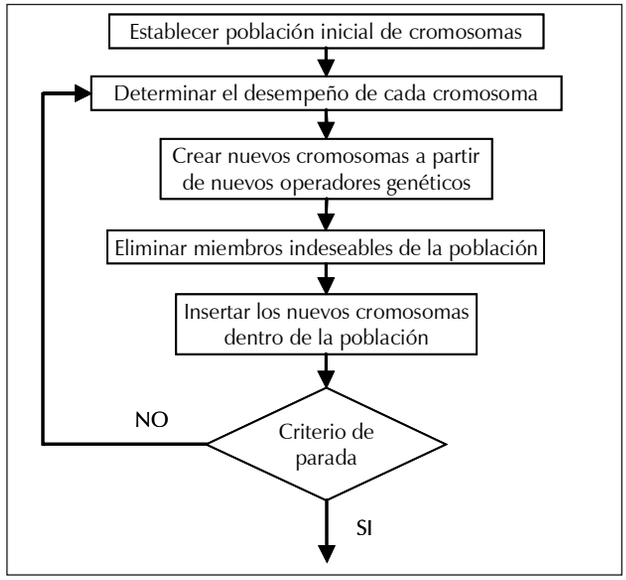


Figura 2. Diagrama de flujo del algoritmo genético

Regla de los k-vecinos más cercanos

Esta es una regla de clasificación simple, por cuanto ofrece un desempeño aceptable con poblaciones con distribución diferente a la normal.

La regla del vecino más cercano puede ser usada como referencia para otros clasificadores, ya que al parecer siempre provee un rendimiento razonable en la mayoría de las aplicaciones (Jain *et al*., 2000). La regla se puede describir como sigue:

1. Definir una medida de distancia entre puntos.
2. Calcular las distancias del punto a clasificar, $\mathbf{x}_0$, a todos los demás puntos de entrenamiento.
3. Seleccionar los *k* puntos muestrales más próximos al que se pretende clasificar.
4. Calcular la proporción en que los *k* puntos pertenecen a cada una de las poblaciones.
5. Clasificar el punto $\mathbf{x}_0$ en la población con mayor frecuencia entre los *k* puntos.

El desarrollo de la regla supone que un punto de observación debe estar situado cerca de los demás puntos muestrales que pertenecen a la misma clase en el espacio de entrenamiento.

Marco experimental

Bases de datos

Señales de voz: se evalúan 90 niños, valorados por el especialista en las clases normal y con hipernasalidad (45 pacientes por cada clase). Cada grabación está conformada por cinco palabras determinadas por fonoaudiólogos que hacen más evidente la enfermedad: */coco/*,*/gato/*,*/jugo/*,*/mano/* y */papá/*. Los registros de las señales fueron tomados en condiciones de bajo nivel de ruido ambiental, usando un micrófono dinámico unidireccional (cardiode), con un rango entre -1 y 1. Todas las señales fueron tomadas en el Hospital Infantil Universitario de Caldas.

Señales de electrocardiografía: los registros corresponden a una base de datos de señales electrocardiográficas (ECG) creada en la Universidad Nacional de Colombia Sede Manizales (BD-ECG-UNCM), durante el período de julio de 2003 y junio de 2004, la cual está diseñada para ser usada en la evaluación de algoritmos entrenados en la detección de cambios de la señal debidos a eventos isquémicos. Las señales usadas para este trabajo corresponden a 50 registros normales y 50 registros que evidencian cardiopatía isquémica. El dispositivo utilizado en la adquisición de los registros en formato digital fue el *Cardio Card PC Based ECG / Resting PC ECG / PC EKG System* fabricado por Nasiff Associates.

Estructura básica de las señales y las características

Los registros de voz son señales no estacionarias que aportan medidas que pueden ser representadas en diferentes dominios (temporal, espectral, cepstral y de modelo inverso). Las características reportan la información acústica del estado funcional (normal o patológico) de acuerdo a niveles previamente establecidos por el especialista. Entre las principales está el *pitch*, una característica que informa sobre la velocidad con que se abren y se cierran las cuerdas vocales; por otro lado, están las características relacionadas con el análisis de los formantes, que son regiones de resonancia o antirresonancia en el espectro de la señal. Es importante anotar que también son consideradas las características de representación abstracta, como son las obtenidas mediante herramientas matemáticas y estadísticas (por ejemplo, los coeficientes Wavelet, las medidas de complejidad, los momentos estadísticos de las variables, etc.). A partir de los registros se obtienen cinco conjuntos de ejemplos (uno por palabra), cada uno consta de 90 observaciones. El número de características por palabra se expone en la Tabla 1.

Tabla 1. Conjunto inicial de características

Palabras	*Número de características*
/coco/	65
/gato/	67
/jugo/	59
/mano/	68
/papá/	43

El electrocardiograma tiene un papel importante en la cardiología, debido a que es un procedimiento efectivo, simple, no invasivo y de bajo costo económico para el diagnóstico de desórdenes cardiovasculares que tienen incidencia epidemiológica alta, generando un impacto relevante en la vida del paciente y en los costos sociales. En la identificación automática de patologías mediante registros de señales electrocardiográficas es importante indicar su amplio carácter no estacionario y cuasiperiódico, donde se requiere detectar formas de onda correspondientes a diferentes estados de normalidad y patología. Durante el entrenamiento del sistema de clasificación, después de las etapas de filtración y segmentación de las señales ECG, se separan 1.800 latidos promediados para el análisis, donde 900 de ellos son considerados normales y los otros restantes con evidencias de cardiopatía isquémica. A partir de esto, se obtiene el conjunto de entrenamiento de 1.010 características correspondiente a los coeficientes Wavelet de las familias *Daubechies*-dbN y *Symlets*-symN de orden 2 a 10, y pares de Wavelets biortogonales como: bior24, bior26, bior28, bior55, bior68, rbio26, rbio28, rbio44, rbio55 y rbio68.

Estructura metodológica para la selección de características

La reducción del espacio de características es llevada a cabo induciendo un árbol de decisión sobre todo el conjunto de entrenamiento, seleccionando sólo las características utilizadas para construirlo. Luego, se desarrolla el método propuesto en (Raymer *et al*., 2000), el cual consiste en encontrar un subconjunto de características de bajo orden y alto poder discriminante, mientras simultáneamente se minimiza el error de clasificación de un conjunto de observaciones dado. Con este fin se utiliza un algoritmo genético, el cual genera y permite la sintonización de los parámetros del método, evaluándolos por el clasificador

de los *k*-vecinos más cercanos (Figura 3). En el algoritmo genético los parámetros del método son representados en un vector de la forma ⟨**w**,*k*,**c**⟩, donde **w** es el vector de longitud *d* que contiene los pesos ⟨w_i:*i*=1,…,*d*⟩ con valores acotados dentro del intervalo [0,1], además, w_i corresponde al peso que escala la *i*-ésima característica, transformando el espacio de características geométricamente de tal forma que se obtenga una mejor separabilidad y, por lo tanto, una mejor clasificación. El número de vecinos más cercanos del clasificador corresponde al valor de *k*. Finalmente, **c** es el vector binario que representa los subconjuntos de características, en donde cada bit está asociado con una; para las *d* características, cada cromosoma tendrá d bits: si el i-ésimo bit es igual a 1, la i-ésima característica es tenida en cuenta para la clasificación. En cambio, si el bit es 0, la característica correspondiente no participa. El procedimiento de validación no es una parte del proceso de selección de características en sí mismo, pero todo método de selección de características debe ser validado (Yu y Liu, 2004). El algoritmo genético implementado en esta metodología tiene las siguientes especificaciones:

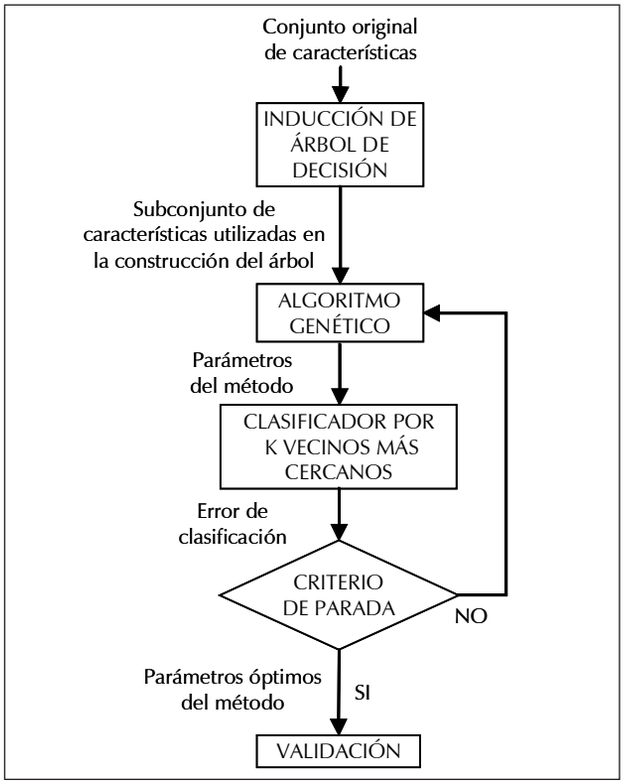
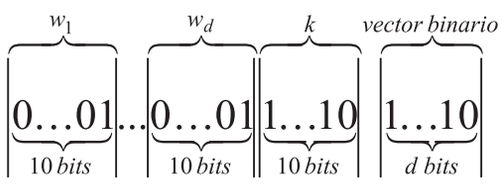


Figura 3. Diagrama de bloques de la metodología propuesta

Codificación. Para las *d* características, los parámetros del método están codificados en cromosomas de longitud 11*d*+10 de la siguiente forma:



Después de realizar un análisis de confiabilidad se estableció el rango de los pesos entre 0 y 1.

Función de evaluación. Dada como el error estimado utilizando el clasificador de los *k*-vecinos más cercanos con el cual se direcciona el algoritmo genético. En las bases de datos de señales de voz, la tasa de error de clasificación se estima utilizando el método de validación cruzada con tres submuestras; en el caso de las bases de datos de electrocardiografía, la tasa de error de clasificación se estima utilizando el 70% del conjunto de ejemplos para entrenamiento y el 30% para validación.

Mutación. Se ha encontrado que variar la tasa de mutación durante la ejecución del AG provee mejores resultados (Eiben *et al.*,1999). Por lo tanto, se utiliza un esquema de control determinístico de la probabilidad de mutación en función de la generación dada por la ecuación:

$$p_m(t) = (1 + (L - 1)\, t/T)^{-1}, \qquad (7)$$

donde *T* es el número total de generaciones, mientras *t* es la generación actual, de tal manera que se cumple $0 \le t \le T$. Esta función disminuye en decrementos de p_m, tal que $p_m(0)=1$ y $p_m(T)=1/T$. La ecuación (7) es una modificación propuesta en (Bäck y Shutz, 1996), con el fin de obtener una mayor diversidad en la población inicial.

Cruzamiento. En la naturaleza, una cría tiene dos padres y hereda los genes de ambos. El operador de cruzamiento se realiza a partir de un par de cromosomas seleccionados con una probabilidad p_C. Se elige de forma aleatoria un punto de cruzamiento, y las cadenas son separadas y cruzadas en el punto. Este proceso de cruzamiento puede ser generalizado a la forma de cruzamiento multipunto; empíricamente se ha demostrado que se puede degradar el desempeño del algoritmo genético a medida que se aumenten los puntos de cruzamiento, ya que se convertiría en una mezcla aleatoria y pocas estructuras importantes de los cromosomas pueden ser preservadas.

Reproducción. Es un proceso en el cual algunos individuos son elegidos como padres para el cruzamiento, de acuerdo a su desempeño dado por la función de evaluación. En este trabajo se utilizó el método de la ruleta, cuyas ranuras tienen un tamaño proporcional al valor del desempeño de los individuos; los individuos con mayor desempeño tendrán una mayor probabilidad de ser seleccionados. Esta técnica se ilustra en los siguientes pasos (Lee, 1996):

- Sumar el valor del desempeño de toda la población y denominar a este resultado como el desempeño total.
- Generar *n*, un número aleatorio entre cero y el desempeño total.
- Retornar el primer miembro de la población cuyo desempeño, sumado al desempeño de los miembros de la población precedentes, es mayor o igual a *n*.

Este esquema de selección tiene el problema potencial de que el mejor miembro de la población puede fallar para

producir una cría en la próxima generación y causar el llamado error estocástico. Para reducir este tipo de error, al procedimiento de la ruleta se le suma una estrategia elitista, la cual consiste en copiar el mejor miembro de cada generación en la siguiente, mejorando el desempeño del algoritmo genético.

Definición de parámetros. En (De Jong, 1975) y (Grefenstette, 1986) se determinan de manera experimental los valores adecuados para los parámetros de un algoritmo genético, de los cuales se eligieron los siguientes: tasa de cruzamiento: 0.95; tamaño de la población: 35; criterio de parada: el algoritmo genético termina cuando se ejecute el número total de generaciones.

Validación

En la práctica, el error de clasificación de un sistema de reconocimiento debe ser estimado a partir de todas las muestras disponibles, las cuales son divididas en conjuntos de entrenamiento y prueba del clasificador. Si el conjunto de entrenamiento es pequeño, el clasificador resultante no será muy robusto y tendrá una baja capacidad de generalización. Por otro lado, si el conjunto de prueba es pequeño, la confiabilidad del error estimado será baja. Los métodos comúnmente usados para estimar la tasa de error son los de *entrenamiento y prueba* y *validación cruzada*. Cuando se tiene un número grande de muestras en comparación con el de características, se puede utilizar el método de entrenamiento y prueba, el cual consiste en tomar de las muestras un conjunto para entrenar el clasificador, usualmente el 70% u 80%, y el resto para la prueba. Cuando el número de muestras disponibles es pequeño en comparación con el de características, es conveniente utilizar el método de validación cruzada, el cual consiste en dividir los datos en *m* submuestras; cada submuestra es clasificada mediante la regla de clasificación construida con las (*m-1*) submuestras restantes, y la tasa de error estimada es la promedio de estas *m* submuestras. El método *leave-one-out* es una validación cruzada con *m* igual al número de ejemplos (Jain *et al*., 2000).

Resultados

La metodología desarrollada se compara con dos métodos de selección de características de amplio uso en la literatura: selección secuencial hacia adelante (SFS) y selección secuencial hacia atrás (SBS), siendo el criterio de evaluación el error de clasificación del subconjunto de características. La comparación se realiza empleando el clasificador del vecino más cercano (1-NN) teniendo como criterio el método de estimación de error *leave-one-out*.

Los resultados experimentales, que se presentan en las tablas 2 a la 6 corresponden a la clasificación de señales de voz. Mientras, en la Tabla 7 se muestran los resultados obtenidos en la identificación de isquemia. Los resultados, en ambos casos, incluyen el tamaño final después de la reducción, N_C, el porcentaje de reducción de características que ofrece el método, r_C, y el error de clasificación, ε, el cual es estimado mediante el método de validación cruzada para el caso de la base de datos de voz, y el método *leave-one-out* para la base de datos de electrocardiografía.

Tabla 2. Resultados de selección para la palabra /coco/

	N_C	r_C(%)	ε (%)
1-NN	65	0,00	28,88
SFS/1 -NN	8	87,69	12,22
SBS/1-NN	11	83,07	16,67
Modelo híbrido	11	83,07	5,55

Tabla 3. Resultados de selección para la palabra /gato/

	N_C	r_C(%)	ε (%)
1-NN	67	0,00	28,88
SFS/1 -NN	29	56,71	16,67
SBS/1 -NN	27	59,70	14,44
Modelo híbrido	5	92,54	8,88

Tabla 4. Resultados de selección para la palabra /jugo/

	N_C	r_C(%)	ε (%)
1-NN	59	0,00	21,11
SFS/1 -NN	18	69,49	16,67
SBS/1-NN	10	83,05	17,78
Modelo híbrido	8	86,44	4,44

Tabla 5. Resultados de selección para la palabra /mano/

	N_C	r_C(%)	ε (%)
1-NN	68	0.00	26.66
SFS/1-NN	5	92.64	15.56
SBS/1-NN	15	83.05	17.78
Modelo híbrido	7	89.70	5.55

Tabla 6. Resultados de selección para la palabra /papá/

	N_C	r_C(%)	ε (%)
1-NN	43	0.00	27.77
SFS/1-NN	8	81.40	20.00
SBS/-NN	11	74.42	24.44
Modelo híbrido	9	79.07	6.66

Tabla 7. Resultados de selección en base de señales de electrocardiografía

	N_C	R_C(%)	ε (%)
1-NN	1010	0.00	2.33
SFS/NN	No converge		
SBS/NN	No converge		
Modelo híbrido	13	98.72	0.39

Conclusiones y trabajo futuro

En este trabajo se desarrolló un modelo híbrido de selección de características con el que se logró reducir la dimensión del espacio de entrenamiento sin comprometer la precisión de clasificación, obteniendo mejores resultados en la reducción de características que con otros esquemas comúnmente usados para la selección de características reportados en la literatura. El modelo incluyó la inducción de un árbol de decisión para la generación de subconjuntos de características, la evaluación de la relevancia mediante el criterio del mínimo error de clasificación y la sintonización del modelo híbrido usando algoritmos genéticos, con lo cual se obtuvo, de forma simultánea, la minimización tanto del número de características de entrenamiento como del error de clasificación.

De manera adicional, a diferencia de las técnicas convencionales de selección, el modelo propuesto informa acerca del nivel de relevancia correspondiente a cada característica perteneciente al conjunto reducido que se obtuvo. Por otro lado, la capacidad de generalización del método propuesto es inducida por el uso de la regla de los *k*-vecinos más cercanos en la función de evaluación.

Debido a que en este trabajo se asoció una sola métrica en la función de evaluación obteniendo resultados aceptables (una tasa promedio de error de clasificación inferior al 6%), como trabajo futuro se sugiere la inclusión de diferentes métricas en la función de evaluación, orientado a la construcción de un sistema de optimización multiobjetivo con ajuste adaptativo.

Agradecimientos

Este trabajo se realizó dentro del marco del proyecto "Auscultación y registro electrocardiográfico sobre la web para apoyo a la teleconsulta médica", código 11271414907, financiado por Colciencias.

Bibliografía

Back, T. y Shutz, M., Intelligent mutation rate control in canonical genetic algorithms: Lecture notes in artificial intelligence., 1996.

Bast, H., Dimension reduction: A powerful principle for automatically finding concepts in unstructured data., In proceedings of the international Workshop on Self-Properties in Complex Information Systems (SELF-STAR'04), 2004, pp 113-116.

Duda, R. O., Hart, P. E. and Store, D. G., Pattern Classification., John Wiley & Sons, 2000.

De Jong, K., The Analysis of the Behaviour of a Class of Genetic Adaptative Systems., Tesis presentada a la Universidad de Michigan, Ann Arbor, para optar por el título de Doctor of Philosophy, 1975.

Eiben, R., Hinterding, and Michalewichz, Z., Parameter control in evolutionary algorithms., IEEE Transactions on Evolutionary Computation, Vol. 3, N° 2, 1999, pp. 124-141.

Grefenstette, J. J., Optimization of control parameters for genetic algorithms., IEEE Transactions on Systems, Man and Cybernetics, Vol. 16, N° 1, 1986, pp. 122-128.

Hong, J. H. and Cho, S. B., Efficient huge-scale feature selection with speciated genetic algorithm., *PRL(27),* No. 2, 15 January, 2006, pp. 143-150.

Jain, A. K., Duin, R. P. W. and Mao, J., Statistical pattern recognition: a review., IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, N° 1, 2000, pp. 4 – 37.

Kim, K. M., Park, J. J., Song, M. H., Kim, I. C., and Suen, C. Y., Binary decision tree using genetic algorithm for recognizing defect patterns of cold mill strip., En Canadian AI 2004, LNAI 3060, A.Y. Tawfik, S. D. Goodwin, editores. Springer-Verlag, Berlín Heidelberg, 2004, pp. 461-466.

Lee, C. S., Neural fuzzy systems: a neuro-fuzzy synergism to intelligent systems., Prentice- Hall, 1996.

Peña, D., Análisis de datos multivariantes., Mc Graw Hill, 2002.

Quinlan, J., Induction of decision trees., Machine Learning, Vol. 1, N° 1, 1986, pp. 81 – 106.

Raymer, M. L., Punch, W. F., Goodman, E. D., Kuhn, L. A. and Jain, A. K., Dimensionality reduction using genetic algorithms., IEEE Transactions on Evolutionary Computation, Vol. 4, N°2, 2000, pp. 164-171.

Yu, L. and Liu, H., Efficient feature selection via analysis of relevance and redundancy., Journal of Machine Learning Research, 5, 2004, pp. 1205–1224.