

# Un estudio sobre el uso de algoritmos genéticos multimodales para selección de características

J.J. Aguilera<sup>1</sup>, M. Chica<sup>1</sup>, M.J. del Jesus<sup>1</sup> y F. Herrera<sup>2</sup>

*Resumen—*

En cualquier proceso de extracción de conocimiento suele ser necesario aplicar una etapa de reducción de la dimensionalidad previa a la aplicación del algoritmo de minería de datos. Una de las formas de llevar a cabo esta tarea es reducir el número de variables o características utilizadas para describir el problema. Esta simplificación permite extraer más información sobre el problema al descartarse variables erróneas (que introducen ruido), variables irrelevantes, y/o variables redundantes (cuya información se puede deducir a través de otras variables). Además, reduce la dimensionalidad del problema a resolver por el algoritmo de minería de datos posterior y facilita la extracción de conocimiento más interpretable y descriptivo, al estar basado en un menor número de características.

En este trabajo se proponen y analizan distintas propuestas de algoritmos genéticos multimodales para selección de características como etapa de reducción de la dimensionalidad previa al diseño de un sistema de clasificación basado en reglas difusas. La experimentación realizada muestra la adecuación de este tipo de algoritmos genéticos para el problema de la selección de características.

*Palabras clave—*Selección de características, optimización multimodal, algoritmos genéticos, fitness sharing, crowding, clearing.

## I. INTRODUCCIÓN

Los Sistemas de Clasificación Basados en Reglas Difusas (SCBRDs) están formados por un conjunto de reglas difusas y un método de inferencia difuso que generaliza el conocimiento extraído de los datos con el fin de clasificar nuevas instancias. Tradicionalmente el principal objetivo en el diseño de este tipo de sistemas ha sido la maximización de la precisión, aunque actualmente está tomando importancia la interpretabilidad del conjunto de reglas difusas generadas [6]. Algunos de los aspectos determinantes de la interpretabilidad de un conjunto de reglas difusas son el tipo y número de reglas, la definición de los conjuntos difusos y el número de variables implicadas en cada una de las reglas.

Cuando el problema de clasificación a resolver tiene un elevado número de variables, el proceso de aprendizaje de un SCBRD interpretable es más difícil por la dimensionalidad de dicho problema, ya que un incremento en el número de variables consideradas implica un incremento exponencial del espacio de búsqueda de reglas difusas.

En la literatura especializada se han considerado dos vías de solución para el problema de la alta dimensionalidad en el diseño de un SCBRD:

- Diseñar una etapa de post-procesamiento en la que se compacte y reduzca un conjunto de reglas obtenido previamente ([1], [11]).
- Realizar un proceso de selección de características que determine las variables más relevantes antes o durante el proceso de aprendizaje inductivo del sistema ([7], [8], [17]). Si el proceso de selección de características se realiza antes del diseño del SCBRD, se reduce el espacio de búsqueda del algoritmo de aprendizaje inductivo y se incrementa su eficiencia y eficacia. En este trabajo se aborda el problema de la alta dimensionalidad desde esta última perspectiva.

Un proceso de selección de características se puede definir como un proceso de búsqueda de un subconjunto del conjunto completo de variables con el objetivo de eliminar características irrelevantes y/o redundantes y obtener un sistema de clasificación más sencillo y con mayor precisión en la clasificación. Es un problema de optimización con restricciones resuelto mediante Algoritmos Genéticos (AGs) en distintas propuestas ([2], [7], [28], [32], [34]).

La selección de características tiene un fuerte carácter multimodal puesto que pueden existir múltiples óptimos (locales o globales) y un AG estándar tiene dificultad para obtener una solución adecuada sin quedar estancado en una solución buena sólo parcialmente. Además, en determinados problemas es conveniente obtener no sólo un conjunto de características adecuado sino los distintos subconjuntos óptimos. En este trabajo se presenta un estudio sobre el uso de distintas propuestas de AGs multimodales [29] para procesos de selección de características previos al diseño de un SCBRD.

Para ello el artículo se organiza de la siguiente forma: en la sección II se introducen algunos conceptos preliminares. En la sección III se presentan las distintas propuestas de AGs multimodales para selección de características y en la sección IV se muestra la experimentación realizada junto con un análisis de los resultados obtenidos. Por último, en la sección V se exponen las conclusiones y se plantean trabajos futuros.

<sup>1</sup>Universidad de Jaén. {jjaguile,mchica,mjjesus}@ujaen.es.

<sup>2</sup>Universidad de Granada. herrera@decsai.ugr.es.

## II. PRELIMINARES

### A. Sistemas de clasificación basados en reglas difusas

Un SCBRD es un sistema de clasificación automático que utiliza reglas difusas como herramienta de representación del conocimiento. En él se pueden distinguir dos componentes principales:

- La base de conocimiento, formada por la base de datos en la que se define la semántica de los conjuntos difusos utilizados y la base de reglas difusas.
- El método de razonamiento difuso, que es el procedimiento de inferencia que combina las reglas con el ejemplo a clasificar y determina la clase a la que pertenece.

En la bibliografía especializada la mayor parte de los SCBRDs utilizan como método de razonamiento difuso el del máximo o el de la suma [12]. En el primero de ellos la clase del nuevo ejemplo es aquella que proporciona la regla que tiene mayor grado de asociación con dicho ejemplo. En el segundo de ellos se tienen en cuenta todas las reglas que emparejan con el ejemplo, clasificando con la clase más votada.

### B. Selección de características

El problema de la selección de características se puede definir como el proceso de búsqueda de  $P$  características de un conjunto inicial  $S$  de  $N$  variables, con  $P \leq N$ , con el objetivo de eliminar características irrelevantes y/o redundantes y obtener un sistema de clasificación más sencillo y con mayor precisión en la clasificación [21], [25].

En un algoritmo de selección de características se distinguen tres componentes:

- Un *algoritmo de búsqueda*, que explora el espacio de subconjuntos de características.
- Una *función de evaluación*, que proporciona una medida de la adecuación del subconjunto de características seleccionado. Según sea esta función los algoritmos de selección de características se pueden clasificar en modelos de filtro, que utilizan medidas de evaluación basadas en medidas de separabilidad de clases: medidas de distancia, de información, de dependencia o de consistencia, entre otras, y modelos de envoltura, que utilizan como función de evaluación una estimación de la precisión en la clasificación alcanzada por un sistema de clasificación diseñado a partir de las variables seleccionadas.
- Una *función de rendimiento*, que valida el subconjunto finalmente seleccionado por el algoritmo de SC.

Se han desarrollado algoritmos genéticos de selección de características tanto desde el enfoque de filtro como desde el de envoltura ([2], [7], [28], [32], [34]). Sus diferencias residen en la definición del esquema de codificación y de la función de adaptación del algoritmo genético.

Las propuestas evolutivas que se presentan y analizan en este trabajo se encuadran dentro del enfoque envolvente, ya que utilizan como parte de la función

de evaluación la estimación de la precisión obtenida por un SCBRD obtenido mediante un método de generación de reglas sencillo, la extensión del método de Wang y Mendel [33] a problemas de clasificación [10].

## III. ALGORITMOS GENÉTICOS MULTIMODALES PARA SELECCIÓN DE CARACTERÍSTICAS

El problema de selección de características tiene un marcado carácter multimodal dado que dentro del espacio de búsqueda pueden existir múltiples óptimos locales o globales por ser conjuntos de variables que permiten distinguir entre clases de igual forma. En este tipo de problemas, un proceso evolutivo estándar puede provocar la convergencia de toda la población a una zona restringida del espacio de búsqueda, abandonando la exploración del resto de óptimos locales. Habitualmente, los AGs estándar no están habilitados para encontrar el máximo global de una función multimodal [16], [26].

Para solucionar este problema, es adecuado el uso de AGs multimodales [26] que se basan en preservar la diversidad de la población mediante el uso de técnicas de nichos [31]. Los AGs multimodales intentan dividir la población en diferentes nichos de forma que soluciones que ocupen diferentes áreas dentro del espacio de búsqueda sean capaces de sobrevivir en la evolución, independientemente de su calidad, para mantener la diversidad necesaria que permita alcanzar los diferentes óptimos de los problemas multimodales.

Los AGs multimodales se clasifican, según la forma en la que se induce la formación de nichos en:

- *Espaciales*, que potencian la formación de nichos en la población de una misma ejecución del AG. Dentro de esta categoría se incluyen

1. Algoritmos de proporción (*sharing*). Se basan en organizar los individuos de la población en nichos, penalizando la calidad de los individuos en función de su cercanía a otras soluciones. Ejemplos de este tipo de AGs son: *fitness sharing* (método de proporción) [16], *continuously updated sharing* (método de proporción con actualización continua) [27], nichos jerárquicos adaptativos [13], y la técnica de identificación de nichos con *fitness sharing* [24], entre otros.

2. Algoritmos de *crowding*, basados en esquemas de reemplazamiento por proximidad. Ejemplos de este tipo de algoritmos son el *crowding* determinístico [26] y el *crowding* multi-nichos [9].

3. Algoritmos de *clearing* que modifican la calidad de los individuos para crear nichos en los que sobrevivan un cierto número de soluciones y el resto desaparezcan. Los métodos basados en *clearing* [30] y los métodos de selección competitiva restringida [22] son ejemplos de este tipo de AGs.

4. Métodos de competición de especies, que crean diferentes especies dentro de la población y hacen que los individuos de las especies evolucionen y luchen por sobrevivir de forma independiente a los individuos del resto de las especies. Es el grupo de

AGs multimodales más reciente y en él se incluyen propuestas como el AG de conservación de especies [23] y los métodos basados en competición jerárquica rápida [19].

- *Temporales*, que potencian la formación de nichos a lo largo de diferentes ejecuciones del AG. Dentro de esta categoría se incluye, entre otros, el método de nichos secuenciales [4].

Todos ellos se han aplicado para resolver distintos problemas multimodales y han proporcionado resultados muy prometedores [29], [31]. Entre ellos, en este trabajo se han seleccionado para desarrollar y analizar las propuestas evolutivas de selección de características los siguientes modelos: *Fitness sharing*, *clearing* y *crowding* determinístico.

En la bibliografía especializada se han utilizado en algunas ocasiones AGs multimodales para selección de características. En particular, se han desarrollado métodos basados en los modelos *crowding* determinístico y *sharing*. En [14] se utiliza un AG con representación binaria para un proceso de selección de características envolvente que utiliza como función fitness una suma ponderada entre precisión (obtenida mediante una red neuronal) y número de características. Incluye un operador de mutación específico para la red neuronal. En [5] se aplica también un AG basado en *crowding* determinístico para selección de características en un problema de clasificación de vinos chilenos con un AG binario con cruce y mutación estándar para dicha representación. Finalmente, en [20] se describe un AG con *fitness sharing*, *clustering* y *random immigrant* aplicado al problema de selección de características, en problemas de extracción de reglas de asociación. Destaca en esta propuesta, el uso de *random immigrant* [3], cuyo objetivo es preservar la diversidad de la población, sustituyendo los individuos de la población que no lleguen a la media de fitness por otros aleatorios cuando el mejor individuo no cambia tras varias generaciones.

#### A. Elementos comunes de las propuestas

En este trabajo se presentan tres propuestas de AGs multimodales para selección de características con el objetivo de mostrar y analizar su comportamiento en distintos problemas. Las propuestas tienen los siguientes elementos comunes:

- Utilizan codificación binaria, con cromosomas de longitud igual al número máximo de variables de cada problema.
- Todas las propuestas se incluyen dentro del enfoque envolvente. Se ha utilizado como medida de evaluación la estimación de la precisión obtenida con el SCBRD obtenido a partir de dichas variables con un método de inducción de reglas difusas sencillo y conocido: la extensión del método de Wang y Mendel [33] a problemas de clasificación [10].
- Para reflejar en la función fitness un equilibrio entre precisión y reducción de la dimensionalidad, la función fitness base es una suma ponderada de la

precisión y la reducción de la dimensionalidad, mediante la siguiente expresión:

$$fitness(z) = \lambda \cdot error(z) + (1 - \lambda) \cdot \frac{caracs(z)}{total} \quad (1)$$

- La mayor parte de los AGs multimodales utilizan una medida de distancia para saber si dos individuos  $i$  y  $j$  pertenecen o no al mismo nicho ( $d(i, j)$ ). Por el tipo de codificación empleada en estas propuestas, la medida de distancia más apropiada es la distancia de *Hamming*.

- Tanto el AG de selección de características basado en *sharing* como el que está basado en *clearing* son AG generacionales y utilizan como operador de selección el torneo binario.

- Todas las propuestas incluyen el operador de cruce simple en un punto y la mutación aleatoria simple.

En las siguientes subsecciones se describen los elementos particulares de cada una de las propuestas.

#### B. AG-Sharing para selección de características

Este algoritmo fue introducido por Holland [18] y mejorado por Goldberg y Richardson [15]. El método clásico de *sharing* se basa en modificar el espacio de búsqueda del algoritmo genético, penalizando soluciones que pertenezcan a zonas de la población con muchos individuos para favorecer la exploración de otras zonas menos pobladas. Se determina el nivel de población de una zona en base al número de soluciones que están a una distancia menor que un determinado radio de nicho ( $\sigma_s$ ).

Para fomentar la evolución de estos nichos independientes se utiliza una función fitness modificada, igual a la función fitness original (1) dividida por el valor del nicho al que pertenece  $m_i$ .

$$f'_i = \frac{f_i}{m_i} \quad (2)$$

donde  $m_i$  se calcula sumando la función *sharing* de todos los miembros de la población:

$$m_i = \sum_{j=1}^L sh(d_{i,j}) \quad (3)$$

representando  $L$  el tamaño de la población y  $d_{i,j}$  la distancia entre los individuo  $i$  y  $j$ . La función *sh()* va a devolver un valor que representa el grado de similitud entre dos individuos.

$$sh(d_{i,j}) = \begin{cases} 1 - (d_{i,j}/\sigma_s)^\alpha & \text{si } d_{i,j} < \sigma_s \\ 0 & \text{en otro caso} \end{cases} \quad (4)$$

$\sigma_s$  denota el umbral de diferencia entre dos individuos para saber si pertenecen o no a un mismo nicho y  $\alpha$  la pendiente de la función *sharing*, cuyo valor más utilizado es 1.

Este método de *sharing* se incluye en el proceso evolutivo, una vez evaluados todos los individuos y antes de aplicar el operador de selección.

La principal desventaja de los metodos basados en el radio del nicho (*sharing* y *clearing*) es la falta de información a priori del valor más recomendable para este parámetro. Esto hace necesaria una experimentación a priori con el objetivo de adaptarlo al tipo y al tamaño de cada problema concreto.

### C. AG-Clearing para selección de características

El algoritmo *clearing* fue propuesto por Petróski en [30] y se basa en la observación de que en la vida natural los recursos disponibles son diferentes entre especies y están limitados. Esto es lo que permite biodiversidad y hace que la rivalidad entre los individuos de diferentes especies disminuya permitiendo que cohabiten en la misma área. En base a esto, un AG de *clearing* limita la presión selectiva modificando la población después de la etapa de evaluación pero antes de realizar la selección, con el objetivo de mantener en cada nicho el individuo o individuos dominantes (con mejor fitness).

El algoritmo básico de *clearing* mantiene el fitness del individuo dominante y elimina el del resto de los individuos de la subpoblación. Pero el algoritmo se puede generalizar de forma que permita incluir varios individuos dentro del mismo nicho. Surge así el concepto de capacidad de un nicho  $\kappa$ , que se define como el máximo número de ganadores que puede haber en el nicho.

El funcionamiento del algoritmo es el siguiente: La población se ordena decrecientemente de acuerdo a su fitness. El primer individuo es el dominante, ya que no existen individuos mejores que él, y se calculará su distancia con los  $n - 1$  restantes. De esta forma se obtienen los individuos que pertenecen a su nicho (utilizando para ello el parámetro  $\sigma$ , que representa el radio del nicho).

Sólo los  $\kappa$  mejores individuos de cada nicho sobrevivirán. Al fitness del resto de los individuos se le asignará el valor 0. El proceso se repetirá pero sólo con los individuos cuyo fitness sea mayor que 0. El resto del proceso evolutivo continúa con esta población.

### D. AG-Crowding para selección de características

Los AGs basados en *crowding* utilizan un esquema de reemplazamiento basado en proximidad para fomentar la formación de nichos.

El AG de *crowding* determinístico en el que se basa nuestra propuesta fue propuesto por Mahfoud [26]. En dicho AG se elimina la selección y la población se divide aleatoriamente en parejas para la reproducción. Cuando se obtienen los descendientes cada uno de ellos compite contra uno de los padres en un torneo que determina qué individuo sobrevivirá en la siguiente generación. La competición se realiza de la siguiente forma: cada padre compite con su hijo más parecido y una vez determinado esto, la supervivencia se determina en base a la calidad (1).

## IV. EXPERIMENTACIÓN Y ANÁLISIS DE RESULTADOS

Para la experimentación se han utilizado tres bases de ejemplos <sup>1</sup> particionadas con el método de estimación de error 10-validación cruzada, y cuyas características principales se muestran a continuación:

- **Base de datos *Ionosphere*.** Recoge información acerca de distintas pruebas realizadas con radares, comprobando si la estructura que se detecta pertenece o no a la capa ionosfera. Tiene 351 instancias y 34 variables más la clase (dos posibles valores nominales que indican si la prueba pertenece o no a la ionosfera).

- **Base de datos *Wisconsin*.** Almacena datos de pacientes con cáncer, del estado americano de Wisconsin. Incluye 570 instancias descritas mediante 30 variables más la clase, la cual posee 2 posibles valores: maligno y benigno.

- **Base de datos *Vehicle*.** Contiene información técnica sobre las características de vehículos. La forman 846 instancias y 18 variables, pertenecientes a una de las 4 posibles clases. El problema consiste en clasificar una silueta dada de un vehículo entre cuatro posibles: *saab*, *van*, *opel* y *bus*. En este artículo se ha trabajado con una transformación de esta base de ejemplos, en la que el número final de clases queda reducido a dos: *van* (23.52 %) y el resto (76.48 %).

Con estas bases de ejemplos se han creado las etiquetas lingüísticas para cada variable realizando una partición uniforme del dominio, con conjuntos difusos triangulares cruzados a nivel 0.5.

En la experimentación realizada se comparan los resultados obtenidos por nuestras propuestas (AG-sharing, AG-clearing y AG-crowding) con los obtenidos con un AG generacional no multimodal con las características indicadas en la Sección III-A. Se compara también con los resultados obtenidos por el algoritmo de aprendizaje utilizado, Wang y Mendel, sin aplicarle ningún algoritmo de selección de características.

El SCBRD de tipo *Wang y Mendel* se construye utilizando los parámetros siguientes:

- Número de etiquetas difusas: 3.
- Tipo de funciones de pertenencia: triangular.
- Grado de compatibilidad: mínimo.
- Método de razonamiento difuso: dos experimentaciones, una con el del máximo y otra con el de la suma.

Los AGs utilizados en la experimentación, junto a sus parámetros, se muestran en la tabla I. Se han utilizado los siguientes parámetros, comunes a todos los algoritmos:

- Tamaño de la población: 100
- Número de evaluaciones: 5000
- Probabilidad de cruce: 0.6
- Probabilidad de mutación: 0.01

<sup>1</sup>Obtenidas del almacén de datos UCI <http://www.ics.uci.edu/~mllearn/MLRepository.html>

- Ponderación del error  $\lambda$  respecto al número de características para el fitness: 0.7 y 1.0

TABLA I  
ALGORITMOS Y PARÁMETROS UTILIZADOS

Método	Parámetros	
<b>AG típico</b>	Algoritmo generacional sin parámetros adicionales	
<i>Clearing</i>	$\sigma = 3, 2$ según problema	$\kappa = 0, 1, 3$
<i>Sharing</i>	$\sigma = 3, 2$ según problema	$\alpha = 1$
<i>Crowding</i>	Algoritmo genético sin parámetros adicionales	

Para todos los algoritmos se han realizado tres ejecuciones con semillas distintas, mostrando en las tablas de resultados las medias y desviaciones típicas de todas las particiones y ejecuciones. Se muestran los resultados obtenidos, con los dos valores del parámetro  $\lambda$ , con los dos métodos de razonamiento difuso, a partir de SCBRD extraídos con el método de Wang y Mendel a partir de subconjuntos de variables obtenidos con el AG típico, con los AGs multimodales y sin realizar reducción de la dimensionalidad en la base de datos. Las tablas II, III, IV, V, VI y VII muestran dichos resultados.

Los datos de la experimentación se dividen en dos bloques: en el primero se ofrece información sobre el mejor individuo de todas las particiones y ejecuciones (precisión y número de características), y en el segundo sobre la media y desviación típica de la precisión y características seleccionadas para cada partición y ejecución.

TABLA II  
RESULTADOS DE CLASIFICACIÓN PARA EL PROBLEMA  
IONOSPHERE CON  $\lambda = 0,7$

	Mejor solución		Precisión		Reducción	
	Prec.	$ S $	$\bar{x}$	$\sigma$	$\bar{x}$	$\sigma$
Método de razonamiento difuso: suma						
AG	91,67	3	83,67	0,06	3,33	0,88
AG-Cl. 0	<b>94,44</b>	<b>2</b>	86,23	0,04	<b>3,07</b>	0,69
AG-Cl. 1	91,67	3	<b>86,24</b>	0,04	3,10	0,61
AG-Cl. 3	94,12	<b>2</b>	85,49	0,04	3,20	0,71
AG-Sh.	94,12	3	80,04	0,07	4,47	1,09
AG-Cr.	91,18	<b>2</b>	82,25	0,05	3,47	0,86
W. & M.	62,97	34	-	-	-	-
Método de razonamiento difuso: máximo						
AG	91,67	4	80,71	0,06	4,30	1,06
AG-Cl. 0	<b>94,29</b>	<b>3</b>	83,73	0,06	<b>3,67</b>	0,71
AG-Cl. 1	91,43	4	83,01	0,05	3,80	0,61
AG-Cl. 3	91,67	4	<b>84,22</b>	0,06	3,80	0,66
AG-Sh.	91,67	<b>3</b>	77,99	0,06	4,67	1,21
AG-Cr.	<b>94,29</b>	<b>3</b>	80,38	0,06	3,77	1,19
W. & M.	65,54	34	-	-	-	-

Los resultados nos muestran que para todos los problemas se logra superar la precisión inicial cuando clasificamos con todas las características. La di-

TABLA III  
RESULTADOS DE CLASIFICACIÓN PARA EL PROBLEMA  
IONOSPHERE CON  $\lambda = 1$

	Mejor solución		Precisión		Reducción	
	Prec.	$ S $	$\bar{x}$	$\sigma$	$\bar{x}$	$\sigma$
Método de razonamiento difuso: suma						
AG	74,29	13	67,05	0,03	18,67	2,81
AG-Cl. 0	77,14	<b>10</b>	66,76	0,04	16,70	3,47
AG-Cl. 1	<b>80</b>	<b>10</b>	67,03	0,05	<b>16,30</b>	3,83
AG-Cl. 3	77,14	12	66,94	0,04	17,40	2,61
AG-Sh.	77,14	12	<b>67,72</b>	0,03	18,13	2,52
AG-Cr.	77,14	14	67,12	0,04	17,83	2,57
W. & M.	62,97	34	-	-	-	-
Método de razonamiento difuso: máximo						
AG	72,22	14	67,44	0,03	19,50	2,90
AG-Cl. 0	<b>74,29</b>	14	67,54	0,03	19,87	3,35
AG-Cl. 1	73,53	<b>12</b>	67,44	0,03	<b>18,10</b>	2,99
AG-Cl. 3	71,43	15	67,34	0,03	19,03	2,65
AG-Sh.	73,53	<b>12</b>	<b>68,02</b>	0,04	19,30	3,47
AG-Cr.	73,53	<b>12</b>	67,74	0,03	19,70	3,19
W. & M.	65,54	34	-	-	-	-

TABLA IV  
RESULTADOS DE CLASIFICACIÓN PARA EL PROBLEMA  
VEHICLE CON  $\lambda = 0,7$

	Mejor solución		Precisión		Reducción	
	Prec.	$ S $	$\bar{x}$	$\sigma$	$\bar{x}$	$\sigma$
Método de razonamiento difuso: suma						
AG	<b>80,00</b>	1	76,97	0,03	1	0
AG-Cl. 0	78,82	1	<b>77,89</b>	0,03	1	0
AG-Cl. 1	<b>80,00</b>	1	77,10	0,03	1	0
AG-Cl. 3	<b>80,00</b>	1	77,18	0,03	1	0
AG-Sh.	<b>80,00</b>	1	77,18	0,03	1	0
AG-Cr.	<b>80,00</b>	1	77,18	0,03	1	0
W. & M.	76,83	18	-	-	-	-
Método de razonamiento difuso: máximo						
AG	<b>82,35</b>	1	78,25	0,02	1	0
AG-Cl. 0	<b>82,35</b>	1	<b>78,79</b>	0,02	1	0
AG-Cl. 1	<b>82,35</b>	1	78,36	0,02	1	0
AG-Cl. 3	<b>82,35</b>	1	78,17	0,02	1	0
AG-Sh.	<b>82,35</b>	1	78,25	0,02	1	0
AG-Cr.	<b>82,35</b>	1	78,25	0,02	1	0
W. & M.	78,24	18	-	-	-	-

ferencia positiva es mayor conforme aumenta la dimensionalidad del problema, por lo que con *Ionosphere* obtenemos la mayor diferencia y con *Vehicle* los resultados son más similares.

Si comparamos los AGs multimodales con el AG generacional simple vemos que normalmente todos los multimodales mejoran sus resultados, existiendo nuevamente, más diferencia en el problema de mayor dimensionalidad, *Ionosphere*.

Examinando únicamente los resultados de los algoritmos multimodales apreciamos que el algoritmo *clearing* con valor  $\kappa = 0$  es el que mejores resultados obtiene. Sólo en alguna combinación de problema y método de inferencia es rebasado por el AG-*clearing* con  $\kappa = 1$  y  $3$  que se presentan como segundos algoritmos en cuanto a rendimiento general.

TABLA V  
RESULTADOS DE CLASIFICACIÓN PARA EL PROBLEMA  
VEHICLE CON  $\lambda = 1$

	Mejor solución		Precisión		Reducción	
	Prec.	$ S $	$\bar{x}$	$\sigma$	$\bar{x}$	$\sigma$
Método de razonamiento difuso: suma						
AG	83,53	<b>3</b>	79,63	0,04	5,90	1,12
AG-Cl. 0	<b>85,88</b>	<b>3</b>	79,70	0,04	<b>4,97</b>	1,67
AG-Cl. 1	83,53	<b>3</b>	79,67	0,04	5,17	1,64
AG-Cl. 3	83,53	<b>3</b>	79,39	0,04	5,10	1,60
AG-Sh.	84,52	<b>3</b>	<b>80,06</b>	0,04	5,80	1,19
AG-Cr.	83,53	<b>3</b>	<b>79,78</b>	0,04	5,67	1,52
W. & M.	76,83	18	-	-	-	-
Método de razonamiento difuso: máximo						
AG	<b>82,35</b>	10	<b>79,23</b>	0,03	11,90	1,32
AG-Cl. 0	<b>82,35</b>	9	79,09	0,03	11,66	1,17
AG-Cl. 1	<b>82,35</b>	10	78,92	0,03	11,80	1,37
AG-Cl. 3	82,14	9	79,20	0,03	11,77	1,38
AG-Sh.	82,14	<b>8</b>	79,04	0,03	11,70	1,24
AG-Cr.	<b>82,35</b>	9	79,04	0,03	<b>11,63</b>	1,50
W. & M.	78,24	18	-	-	-	-

TABLA VI  
RESULTADOS DE CLASIFICACIÓN PARA EL PROBLEMA  
WISCONSIN CON  $\lambda = 0,7$

	Mejor solución		Precisión		Reducción	
	Prec.	$ S $	$\bar{x}$	$\sigma$	$\bar{x}$	$\sigma$
Método de razonamiento difuso: suma						
AG	96,49	3	<b>92,96</b>	0,03	2,70	0,70
AG-Cl. 0	96,49	2	92,21	0,03	2,53	0,86
AG-Cl. 1	96,49	2	92,33	0,02	2,40	0,77
AG-Cl. 3	96,49	2	92,33	0,03	2,47	0,78
AG-Sh.	<b>98,25</b>	3	92,74	0,03	2,70	0,65
AG-Cr.	94,74	2	92,62	0,02	<b>2,20</b>	0,55
W. & M.	89,09	30	-	-	-	-
Método de razonamiento difuso: máximo						
AG	98,25	1	91,39	0,04	1,73	0,91
AG-Cl. 0	98,25	1	91,39	0,04	1,67	0,71
AG-Cl. 1	98,25	1	91,51	0,04	1,63	0,72
AG-Cl. 3	98,25	1	<b>91,62</b>	0,04	1,67	0,80
AG-Sh.	<b>100</b>	2	91,04	0,04	<b>1,13</b>	0,35
AG-Cr.	<b>100</b>	2	90,57	0,05	1,67	0,48
W. & M.	88,57	30	-	-	-	-

Podemos indicar que para el problema de la selección de características, el método de *clearing* muestra mejores resultados que el *sharing*, *crowding* determinístico, y por supuesto, que un AG simple.

Si observamos las diferencias entre el uso de uno u otro método de inferencia, en este caso, suma o máximo, no podemos afirmar que uno sea mejor que otro, ya que según el problema que examinemos vemos que un método funciona mejor que otro. Así, para *Ionosphere* el método de la suma obtiene mejor precisión, pero para *Vehicle* obtenemos más precisión si usamos como método de inferencia el del máximo. Con *Wisconsin* el método de la suma obtiene más precisión, pero reduce menos características que el método del máximo. Dependiendo del problema será más conveniente usar un método u otro.

TABLA VII  
RESULTADOS DE CLASIFICACIÓN PARA EL PROBLEMA  
WISCONSIN CON  $\lambda = 1$

	Mejor solución		Precisión		Reducción	
	Prec.	$ S $	$\bar{x}$	$\sigma$	$\bar{x}$	$\sigma$
Método de razonamiento difuso: suma						
AG	<b>100</b>	11	95,08	0,03	12,23	0,90
AG-Cl. 0	<b>100</b>	<b>10</b>	<b>95,25</b>	0,03	<b>11,63</b>	1,13
AG-Cl. 1	<b>100</b>	<b>10</b>	94,90	0,03	12,03	1,22
AG-Cl. 3	<b>100</b>	<b>10</b>	94,55	0,03	11,93	1,23
AG-Sh.	<b>100</b>	<b>10</b>	94,32	0,03	12,30	1,42
AG-Cr.	<b>100</b>	11	94,20	0,03	12,27	0,83
W. & M.	89,09	30	-	-	-	-
Método de razonamiento difuso: máximo						
AG	94,74	5	<b>92,85</b>	0,03	<b>9,13</b>	5,02
AG-Cl. 0	<b>100</b>	<b>4</b>	90,10	0,04	11,67	4,31
AG-Cl. 1	<b>100</b>	5	89,22	0,04	11,10	5,40
AG-Cl. 3	<b>100</b>	5	90,21	0,04	11,33	3,63
AG-Sh.	<b>100</b>	7	89,28	0,04	14,60	3,89
AG-Cr.	<b>100</b>	9	89,34	0,04	14,13	4,21
W. & M.	88,57	30	-	-	-	-

Se aprecia también cómo se reducen muchas características en los tres problemas. Tanto es así que para *Vehicle* siempre se considera una única característica (partíamos de 18 características iniciales).

La reducción de características es bastante alta. Esta reducción de la dimensionalidad es más evidente cuando se utiliza como ponderación del error el valor 0,7. Esta combinación de pesos para la función *fitness* potencia la reducción de la dimensionalidad con valores de precisión adecuados. Incluso en problemas con un alto número de variables, como es el caso de *Ionosphere*, incrementa la precisión.

En la tabla VIII se recogen el número medio de evaluaciones necesarios para obtener el mejor individuo en cada una de las propuestas. Como se comentó anteriormente, el número de evaluaciones ha sido 5000 para todas las bases de ejemplos. De los resultados obtenidos se puede apreciar, como cabía esperar, que cuanto menor es el tamaño de los datos, la mejor solución se obtiene antes. Es destacable el hecho de que son los algoritmos AG-Sharing y AG-Crowding los que necesitan un mayor número de evaluaciones para converger en todos los problemas. Esto, unido a los valores obtenidos en precisión y en reducción hace que se pueda considerar, que con la experimentación realizada el algoritmo con un mejor comportamiento global tanto en eficacia como en eficiencia es el AG-clearing.

## V. CONCLUSIONES Y TRABAJOS FUTUROS

En este trabajo hemos descrito y analizado distintas propuestas evolutivas multimodales y un algoritmo genético simple aplicados al problema de la selección de características para tres problemas distintos: *Ionosphere*, *Wisconsin* y *Vehicle*.

Los algoritmos genéticos desarrollados han sido algoritmos de selección de características de tipo en-

TABLA VIII  
 NÚMERO MEDIO DE EVALUACIONES NECESARIOS PARA  
 OBTENER EL MEJOR INDIVIDUO

	Ionosfera	Vehicle	Wisconsin
Ponderación $\lambda$ del fitness = 0.7			
Método de razonamiento difuso: suma			
AG	2360,40	848,92	1651,60
AG-Cl. 0	3117,63	664,92	1864,43
AG-Cl. 1	3020,13	605,07	1963,50
AG-Cl. 3	3018,20	584,62	2076,80
AG-Sh.	3388,97	1507,03	2505,10
AG-Cr.	4095,13	1361,00	3190,27
Método de razonamiento difuso: máximo			
AG	2256,77	906,80	1927,93
AG-Cl. 0	3658,10	657,93	1927,90
AG-Cl. 1	3140,07	593,63	1765,43
AG-Cl. 3	3214,43	595,23	2049,27
AG-Sh.	3370,67	972,50	3201,07
AG-Cr.	4073,87	1278,30	3654,90

	Ponderación $\lambda$ del fitness = 1		
Método de razonamiento difuso: suma			
AG	1988,57	1353,47	1877,2
AG-Cl. 0	2978,83	2613,7	2496,8
AG-Cl. 1	2734,83	2001,57	2290,97
AG-Cl. 3	2106,03	1851,3	2267,33
AG-Sh.	2061,2	2690,63	2264,63
AG-Cr.	3620,93	3445,37	3735,57
Método de razonamiento difuso: máximo			
AG	1712,5	827,87	1279,83
AG-Cl. 0	2187,13	749,72	2973,87
AG-Cl. 1	2180,77	878,77	3234,3
AG-Cl. 3	1753,47	696,43	2907,03
AG-Sh.	1830,13	1132,83	2310,73
AG-Cr.	2819,13	1560,83	3303,57

voltura basándose en el algoritmo de generación de reglas difusas de *Wang y Mendel*.

Los tres algoritmos multimodales implementados, *clearing*, *sharing* y *crowding determinístico*, han demostrado ser mejores que el algoritmo genético simple y sobre todo, han mejorado ampliamente la precisión inicial con todas las características. Se han seleccionado muy pocas variables, del orden de 3 ó 4 para problemas que inicialmente poseían 34, y aún así se ha mejorado la precisión.

El algoritmo genético con *clearing* se ha probado para distintos valores de  $\kappa$  y se ha visto que es el algoritmo multimodal que mejores prestaciones presenta, tanto en precisión como en reducción como rapidez de convergencia.

La mejor precisión se ha obtenido en *Wisconsin* con un 92.3 % de media y una selección media de 2.4 características de las 30 iniciales.

Los trabajos futuros que nos proponemos desarrollar giran en torno a las siguientes líneas:

- Un estudio y modificación de los algoritmos genéticos multimodales para no quedarnos sólo con el mejor individuo de la población, sino considerar todos los óptimos encontrados: los mejores de cada nicho. Así podríamos extraer más conocimiento sobre el problema analizando relaciones entre las variables; posibilitaríamos que el experto eligiese el sub-

conjunto más adecuado en base a criterios como coste de extracción de las variables, interpretabilidad, etc., y podríamos abordar el diseño de multclasificadores a partir de los subconjuntos obtenidos. Se abordaría también un estudio sobre la diversidad de los resultados obtenidos.

- Una aplicación de estos algoritmos con nichos a problemas de más alta dimensionalidad, sobre los que creemos que obtendrán un rendimiento mucho mayor al tener un campo de soluciones más amplio.
- Desarrollo de nuevas propuestas basadas en AGs multimodales avanzados.

#### AGRADECIMIENTOS

Este trabajo está soportado por los proyectos TIN 2005-08386-C05-03 y TIN 2005-08386-C05-01.

#### REFERENCIAS

- [1] J. Abonyi, J.A. Roubos, F. Szeifert, *Data-driven generation of compact, accurate, and linguistically sound fuzzy classifiers based on a decision-tree initialization*, International Journal of Approximate Reasoning 32 (1): 1-21, 2003.
- [2] J. Bala, K. De Jong, J. Huang, H. Vafaie, H. Wechsler, *Using learning to facilitate the evolution of features for recognizing visual concepts*, Evolutionary Computation 4(3): 297-311, 1997.
- [3] C. Bates Congdon, *A comparison of genetic algorithms and other machine learning systems on a complex classification task from common disease research*, PhD thesis, University of Michigan, 1995.
- [4] D. Beasley, D.R. Bull, R.R. Martin *A sequential niche technique for multimodal function optimization* Evolutionary Computation 1(2): 101-125, 1993.
- [5] N.H. Beltrán, M.A. Duarte-Mermoud, S.A. Salah, M.A. Bustos, A.I. Peña-Neira, E.A. Loyola, J.W. Jalocho *Feature selection algorithms using Chilean wine chromatograms as examples*, Journal of food engineering, 67: 483-490, 2005.
- [6] J. Casillas, O. Cordón, F. Herrera, L. Magdalena (Eds.), *Interpretability issues in fuzzy modeling*, Springer-Verlag Series Studies in Fuzziness and Soft Computing 128, 2003.
- [7] J. Casillas, O. Cordón, M.J. Del Jesus, F. Herrera, *Genetic feature selection in a fuzzy rule-based classification system learning process for high-dimensional problems*, Information Sciences 136: 135-157, 2001.
- [8] G. Castellano, C. Castiello, A.M. Fanelli, C. Mencar, *Knowledge discovery by a neuro-fuzzy modeling framework*, Fuzzy Sets and systems 149 (1): 187-207, 2005.
- [9] W. Cedeño, V.R. Vemuri, *Analysis of speciation and niching in the multiniche crowding GA*, Theoretical Computers Science 229: 177-197, Elsevier, 1999.
- [10] Z. Chi, H. Yan, T. Pham, *Fuzzy algorithms with applications to image processing and pattern recognition*, World Scientific, Singapore, 1996.
- [11] O. Cordón, M.J. Del Jesus, F. Herrera, M. Lozano, *MO-GUL: a methodology to obtain genetic fuzzy rule-based systems under the iterative rule learning approach*, International Journal of Intelligent Systems 14 (11): 1123-1143, 1999.
- [12] O. Cordón, M.J. Del Jesus, F. Herrera, *A proposal on reasoning methods in fuzzy rule-based classification systems*, International Journal of Approximate Reasoning 20(1): 21-45, 1999.
- [13] G. Dunwei, P. Fengping, X. Shifan, *Adaptive niche hierarchy genetic algorithm*, Actas de IEEE TENCON: 39-42, 2002.
- [14] P.A. Estévez, M. Fernández, R.J. Alcock, M.S. Packianather, *Selection of Features for the Classification of Wood Board Defects*, IEEE on Artificial Neural Networks, : 347-352, 7-10 septiembre 1999.
- [15] D.E. Goldberg y J. Richardson, *Genetic Algorithms with sharing for multimodal function optimization.*, In proceeding of the Second International Conference on Gene-

- tic Algorithms: 41-49. J.J. Grefenstette Ed. Hillsdale, NJ: Lawrence Erlbaum. 1987.
- [16] D.E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning.*, Addison-Wesley 1989.
  - [17] A. González, R. Pérez, *Selection of relevant features in a fuzzy genetic learning algorithm*, IEEE Transactions on Systems, Man and Cybernetics Part B 31(3): 417-425, 2001.
  - [18] J.H. Holland, *Adaptation in Natural and Artificial Systems.*, Ann Arbor, MI: University of Michigan Press, 1975.
  - [19] J.J. Hu, E.D. Goodman, *Robust and efficient genetic algorithms with hierarchical niching and a sustainable evolutionary computation model*, Proc. GECCO: 1220-1232, 2004.
  - [20] L. Jourdan, C. Dhaenens y El-Ghazali Talbi, *A Genetic Algorithm for Feature Selection in Data-Mining for Genetics*, MIC'2001 - 4th Metaheuristics International Conference: 29-34, Portugal, 2001.
  - [21] Kohavi, R., John, G.H.: *Wrappers for Feature Subset Selection*, Artificial Intelligence 97: 273-324, 1997
  - [22] Ch. Lee, D. Cho, H. Jung, *Niching genetic algorithm with restricted competition selection for multimodal function optimization*, IEEE Transactions on Magnetics 35:1722-1725, 1999.
  - [23] J. Li, M. Balazs, G.T. Parks, P.J. Clarkson, *A species conserving genetic algorithm for multimodal function optimization*, Evolutionary computation 10(3): 207-234, 2002.
  - [24] C. Lin, W. Wu, *Niche identification techniques in multimodal genetic search with sharing scheme*, Advances in Engineering Software (33): 779-791. Elsevier, 2002.
  - [25] H. Liu y H. Motoda, *Feature Selection for Knowledge Discovery and Data Mining*, Editorial Kluwer Academic Publishers, 1998.
  - [26] S. W. Mahfoud, *Niching methods for Genetic Algorithms*, Ph.D. Thesis, University of Illinois, Urbana-Champaign, 1995.
  - [27] C.K. Oei, D.E. Goldberg y S.J. Chang *Tournament selection, niching and the preservation of diversity*, Informe Técnico 91011. Illinois Genetic Algorithms Laboratory. Universidad de Illinois, Urbana, 1991.
  - [28] I.S. Oh, J.S. Lee, B.R. Moon, *Hybrid Genetic Algorithms for Feature Selection*, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (11): 1424-1437, 2004.
  - [29] E. Pérez, F. Herrera y C. Hernández, *Finding multiple solutions in job shop scheduling by niching genetic algorithms*, Journal of Intelligent Manufacturing, 14: 323-341, 2003.
  - [30] A. Pérowski, *Clearing procedure as a niching method for genetic algorithms*, In Proceedings of the IEEE Third International Conference on Evolutionary Computation: 798-803, Nagoya , Japan, 1996.
  - [31] B. Sareni y L. Krahenbuhl, *Fitness sharing and niching methods revised*, IEEE Transactions on Evolutionary Computation, 2: 97-106, 1998.
  - [32] W. Siedlecki, J. Sklansky, *A note on genetic algorithm for large-scale feature selection*, Pattern Recognition Letters 10: 335-347, 1989.
  - [33] L.X. Wang y J.M. Mendel, *Generating fuzzy rules by learning from examples*, IEEE Transactions Systems Man Cybern., 25: 353-361, 1992.
  - [34] P. Zhang, B. Verma, K. Kumar, *Neural vs. statistical classifier in conjunction with genetic algorithm based feature selection*, Pattern Recognition Letters 26(7): 909-919, 2005.