

# Resolución de Problemas Multi-clase con Sistemas de Clasificación Basados en Reglas Difusas Lingüísticos Basados en Aprendizaje por Parejas y Relaciones de Preferencia

Alberto Fernández<sup>1</sup> María Calderón<sup>2</sup> Edurne Barrenechea<sup>2</sup> Humberto Bustince<sup>2</sup> Francisco Herrera<sup>1</sup>

<sup>1</sup> Dept. de Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada, España, {alberto,herrera}@decsai.ugr.es

<sup>2</sup> Dept. de Automática y Computación, Universidad Pública de Navarra, España, Maria.CalderonTurullols@cajanavarra.es, {edurne.barrenechea, bustince}@unavarra.es

## Resumen

Esta contribución versa sobre el uso de Sistemas de Clasificación Basados en Reglas Difusas Lingüísticas en problemas con múltiples clases. La idea es decomponer el conjunto de datos original en problemas de clasificación binarios usando la metodología de aprendizaje por parejas (enfrentando todas las parejas de clases), y obtener un sistema difuso independiente para cada una de ellos. A lo largo del proceso de inferencia, cada subclasificador genera un grado de asociación para sus dos clases y estos valores se codifican en una relación de preferencia difusa.

Nuestro análisis se centra en el paso de predicción de la clase. Concretamente, proponemos el manejo de la relación de preferencia difusa usando un criterio de no dominancia sobre las diferentes alternativas, contrastando el comportamiento de este modelo con una estrategia de voto. El objetivo es mostrar la bondad del uso del aprendizaje por parejas y estudiar cómo afecta cada una de las alternativas de clasificación seleccionadas en el rendimiento final del modelo.

**Palabras Clave:** Sistemas de clasificación basados en reglas difusas, problemas multi-clase, relaciones de preferencia difusas, multi-clasificadores, aprendizaje por parejas.

## 1 Introducción

Los Sistemas de Clasificación Basados en Reglas Difusas (SCBRDs)[12] son una herramienta popular en Aprendizaje Automático debido a la ventaja del uso de términos lingüísticos para incrementar la interpretabilidad global del modelo de salida.

En este trabajo, centramos nuestra atención en el problema de múltiples clases, dado que este tipo de datos son muy comunes en aplicaciones del mundo real y pueden representar una dificultad añadida para los clasificadores difusos para encontrar una correcta separabilidad en un espacio de búsqueda con un alto número de clases. Para simplificar el problema original, proponemos el uso del enfoque de aprendizaje por parejas [8], que consiste en entrenar un clasificador para cada posible par de clases ignorando los ejemplos que no pertenecen a las clases relacionadas. Para determinar la salida entre todas las predicciones para los clasificadores asociados, los representaremos en una relación de preferencia difusa transformando el problema de clasificación en un problema de toma de decisión. En este trabajo, proponemos el uso de un criterio de no dominancia maximal [16] para el proceso de decisión final, dado que hemos mostrado previamente su buen comportamiento en sistemas de clasificación lingüísticos [4]. El marco conceptual de la clasificación basada en relaciones de preferencia difusas fue propuesto por Hüllermeier y Brinker en [11]. Adicionalmente, Hühn y Hüllermeier propusieron en [10] el algoritmo FR3, un SCBRD aproximativo que usa una regla de decisión basada en una estrategia de voto para tomar la decisión final en el proceso de clasificación.

En nuestro estudio, determinaremos la bondad del enfoque de aprendizaje por parejas para sistemas difusos lingüísticos analizando las diferencias en rendimiento obtenidas por el modelo clásico de SCBRD y el enfoque de multi-clasificación utilizando un método de aprendizaje de reglas difusas con un buen comportamiento, el algoritmo “Fuzzy Hybrid Genetics-Based Machine Learning” (FH-GBML) [14, 15]. Además, analizaremos el mencionado criterio de no dominancia que proponemos en contraste con la estrategia de voto introducida por Hühn y Hüllermeier en [10], para tener conclusiones sobre la utilidad del aprendizaje por parejas para SCBRD lingüísticas.

Hemos seleccionado 14 conjuntos de datos multiclase del repositorio UCI [2] dentro del marco experimental. La medida de rendimiento se basa en el porcentaje de acierto y la calidad de los resultados está apoyada por un correcto análisis estadístico tal y como se sugiere en la literatura [3, 7].

Para ello, este trabajo se organiza como sigue. En la Sección 2 presentamos el concepto de multi-clasificación, e introducimos nuestra propuesta de aprendizaje por parejas usando relaciones de preferencia, también describiendo el algoritmo FH-GBML seleccionado para nuestro estudio. La Sección 3 incluye el marco experimental, esto es, los conjuntos de datos para el testeo, los parámetros de configuración y los tests estadísticos para la comparativa del rendimiento. En la Sección 4 presentamos nuestro análisis empírico. La Sección 5 concluye el trabajo.

## 2 Sistemas de Multi-Clasificación Basados en Reglas Difusas

En este trabajo, buscamos mejorar el rendimiento de los SCBRDs introduciendo una propuesta de multi-clasificación basada en aprendizaje por parejas [5, 8]. Esta metodología permite una mejor exploración del dominio del problema dividiendo el conjunto de datos original en subproblemas binarios que son más fáciles de discriminar.

En esta sección introduciremos primero el concepto de multi-clasificación y el esquema de aprendizaje para un sistema difuso lingüístico basado en aprendizaje por parejas. A continuación, definiremos el procedimiento usado para obtener la clasificación final entre todos los clasificadores por medio de las relaciones de preferencia difusas. Finalmente, describiremos el algoritmo FH-GML, que será empleado como la base del modelo difuso.

### 2.1 Problemas Multi-Clase Via Aprendizaje por Parejas

Existe una gran cantidad de aplicaciones que requieren una categorización multi-clase. Para simplificar el proceso de clasificación, podemos transformar las fronteras del problema distinguiendo sólo entre dos clases dividiendo el problema inicial en múltiples conjuntos de dos clases que pueden ser resueltos por separado.

Concretamente, hemos considerado el enfoque de aprendizaje por parejas [8], que consiste en entrenar un clasificador para cada posible pareja de clases ignorando los ejemplos que no pertenecen a las clases asociadas. En el momento de la clasificación, se somete una instancia de consulta a todos los modelos binarios, y las predicciones de estos modelos se com-

binan en una clasificación conjunta.

Cada uno de estos clasificadores difusos tendrá su propia Base de Conocimiento (BC), compuesta por una Base de Datos (BD) y una Base de Reglas (BR). Hemos decidido centrarnos en la precisión del modelo contextualizando las particiones difusas para cada subproblema por separado, tomando como universo de discurso el rango de los valores de las variables del subconjunto de ejemplos para cada subconjunto de entrenamiento.

La BR para cada clasificador se aprende usando un método de aprendizaje de reglas, que puede seleccionarse entre las diferentes propuestas de la literatura especializada.

Una vez que todas las BCs han sido aprendidas, procedemos con el paso final de inferencia. Cuando un nuevo patrón de entrada se presenta al sistema, cada SCBRD es disparado para definir el grado de salida para su pareja de clases asociadas.

### 2.2 Proceso de Clasificación con Relaciones de Preferencia Difusas Basado en No-Dominancia

Consideraremos el problema de clasificación como un problema de toma de decisión, y definiremos una relación de preferencia difusa  $R^b$  [16] con las salidas correspondientes de los SCBRDs. De este modo, el cálculo de cada grado de preferencia se basa en la función de agregación que combina los grados de asociación positivos entre las reglas difusas y el patrón de entrada, lo que se conoce como Método de Razonamiento Difuso.

$$R^b = \begin{bmatrix} - & r_{1,2} & \cdots & r_{1,m} \\ r_{2,1} & - & \cdots & r_{2,m} \\ \vdots & \ddots & \ddots & \vdots \\ r_{m,1} & r_{m,1} & \cdots & - \end{bmatrix} \quad (1)$$

Consideramos el emparejamiento máximo, donde cada nuevo patrón  $x_p$  se clasifica como la clase consecuente de una única regla ganadora ( $Class(x_p) = C_w$ ) que se determina como

$$\mu_{A_w}(x_p) \cdot RW_w = \max\{\mu_{A_q}(x_p) \cdot RW_q, Regla_q \in BR\} \quad (2)$$

donde  $\mu_{A_q}(x_p)$  es el grado de pertenencia del ejemplo patrón  $x_p = (x_{p1}, \dots, x_{pn})$  con el antecedente de la regla  $R_q$  y  $RW_q$  es el peso de la regla.

Por lo tanto,  $R^b(i, j)$  (el grado de preferencia difuso entre las clases  $i$  y  $j$ ) es el máximo grado de asociación para todas las reglas en la BR que concluyen la clase  $i$ .  $R^b(i, j)$  será normalizado a  $[0, 1]$  mediante la expresión (3), obteniendo la relación  $R(i, j) = 1 - R(j, i)$ .

$$R(i, j) = \frac{R^b(i, j)}{R^b(i, j) + R^b(j, i)} \quad (3)$$

A partir de la relación de preferencia difusa podemos extraer un conjunto de alternativas (clases) no dominadas como solución del problema de toma de decisión difuso y de este modo, nuestra salida de clasificación. Concretamente, los elementos maximales no dominados de  $R$  se calculan por medio de las siguientes operaciones, de acuerdo al criterio de no dominancia propuesto por Orlovsky [16]:

- Primero, calculamos la relación estricta de preferencia difusa  $R'$  que es igual a :

$$R'(i, j) = \begin{cases} R(i, j) - R(j, i), & \text{cuando } R(i, j) > R(j, i), \\ 0, & \text{en otro caso.} \end{cases} \quad (4)$$

- Entonces, calculamos el grado de no dominancia de cada clase  $ND_i$ , que se obtiene simplemente como:

$$ND_i = 1 - \sup_{j \in C} [R'(j, i)] \quad (5)$$

Este valor representa el grado en el que la clase  $i$  no es dominada por ninguna de las clases restantes.  $C$  representa el conjunto total de clases en el conjunto de datos. La clase de salida se calcula como el índice del valor maximal de no dominancia:

$$Clase(x_p) = \arg \max_{i=1, \dots, m} \{ND_i\} \quad (6)$$

### 2.3 Algoritmo de Generación de Reglas “Fuzzy Hybrid Genetics Based Machine Learning”

El método FH-GBML [14] consiste en un enfoque Pittsburgh donde cada conjunto de reglas se maneja como un individuo. También contiene un enfoque de aprendizaje genético Cooperativo-Competitivo (un individuo representa una única regla), que se usa como un tipo de mutación heurística para modificar parcialmente cada conjunto de reglas.

Este método utilizar reglas difusas estándar con pesos en las reglas [13] donde cada variable de entrada  $x_i$  se representa mediante un término lingüístico o etiqueta. El sistema define 14 posibles términos lingüísticos para cada atributo así como un conjunto especial “do not care”.

En el proceso de aprendizaje,  $N_{pop}$  conjuntos de reglas se crean seleccionando aleatoriamente  $N_{rule}$  patrones de entrenamiento. Entonces, se genera una regla difusa para cada uno de los patrones de entrenamiento seleccionadas eligiendo probabilísticamente un

conjunto difuso antecedente a partir de los 14 candidatos ( $P(B_k) = \frac{\mu_{B_k}(x_{pi})}{\sum_{j=1}^{14} \mu_{B_j}(x_{pi})}$ ) y cada conjunto difuso antecedente de la regla difusa generada se reemplaza con *don't care* utilizando una probabilidad predeterminada  $P_{don't care}$ .

$N_{pop} - 1$  conjuntos de reglas son generados mediante selección, cruce y mutación en el mismo modo que el algoritmo estilo Pittsburgh. A continuación, con una probabilidad predeterminada, se aplica una única iteración del algoritmo genético Cooperativo-Competitivo para cada uno de los conjuntos de reglas generados.

Finalmente, el mejor conjunto de reglas se añade de la población actual a los nuevos ( $N_{pop} - 1$ ) conjuntos de reglas generados para formar la siguiente población y, si el criterio de parada no se ha satisfecho aún, el proceso genético se repite de nuevo. La clasificación se realiza siguiendo el Método de Razonamiento Difuso del emparejamiento máximo (regla ganadora).

Para más detalles acerca de esta propuesta, por favor refiérase a [14]. Este algoritmo está disponible dentro de la herramienta software KEEL [1] (<http://www.keel.es>).

## 3 Marco Experimental

En esta sección, primero daremos detalles de los problemas multi-clase reales escogidos para la experimentación y los parámetros de configuración de los SCBRDs (subsecciones 3.1 y 3.2 respectivamente). Finalmente, introduciremos los tests estadísticos aplicados para comparar los resultados obtenidos a lo largo del estudio experimental (subsección 3.3).

### 3.1 Conjuntos de datos

La Tabla 1 resume las propiedades de los conjuntos de datos seleccionados. Muestra, para cada conjunto de datos, el número de ejemplos (#Ej.), el número de atributos (#Atrs.) y el número de clases (#Cl.). Los conjuntos de datos *letter*, *penbased* y *page-blocks* han sido muestreados mediante estratificación al 10% para reducir su tamaño para el entrenamiento. En el caso de los valores perdidos (*cleveland*) hemos eliminado esas instancias del conjunto de datos.

La estimación del ratio de acierto fue obtenida por medio de una validación cruzada de 5 partes. Esto es, dividimos el conjunto de datos en 5 partes, cada una conteniendo el 20 % de los patrones del conjunto de datos. Para cada parte, el algoritmo fue entrenado con los ejemplos de las particiones restantes y entonces, testeado con la partición actual.

Tabla 1: Descripción resumen de los conjuntos de datos

id	Conjunto	#Eh.	#Atrs.	#Cl.
bal	balance scale	625	4	3
cle	cleveland	297	13	5
eco	ecoli	336	7	8
gla	glass identification	214	9	7
iri	iris	150	4	3
let	letter	2000	16	26
new	new-thyroid	215	5	3
pag	page-blocks	548	10	5
pen	pen-based recognition	1,099	16	10
seg	segment	2310	19	7
shu	shuttle	2,175	9	5
veh	vehicle	846	18	4
win	wine	178	13	3
yea	yeast	1,484	8	10

### 3.2 Parámetros

La configuración seleccionada para el algoritmo FH-GBML consiste en la T-norma producto como operador de conjunción, junto con el Factor de Certeza Penalizado [13] para el peso de la regla y MRD de la regla ganadora. Respecto a los parámetros específicos para el proceso genético, hemos escogido los siguientes valores:

- Número de reglas difusas:  $5 \cdot d$  reglas.
- Número de conjuntos de reglas: 200 conjuntos.
- Probabilidad de Cruce: 0,9.
- Probabilidad de Mutación:  $1/d$ .
- Número de reglas reemplazadas: Todas las reglas excepto la mejor (parte Pittsburgh), número de reglas / 5 (parte Cooperativa-Competitiva).
- Número total de generaciones: 1.000.
- Probabilidad de “Don’t care”: 0,5.
- Probabilidad de la aplicación de la iteración Cooperativa-Competitiva: 0,5.

donde  $d$  representa la dimensionalidad del problema (número de variables).

### 3.3 Tests estadísticos para la comparativa del rendimiento

En esta contribución, usamos las técnicas de comprobación de hipótesis para proporcionar un soporte estadístico al análisis de los resultados [6, 17]. Concretamente, usaremos tests no paramétrico, dado que las condiciones iniciales que garantizan la fiabilidad de los tests paramétricos puede no satisfacerse, haciendo que el análisis estadísticos pierda credibilidad con estos tipos de tests [3, 7]. Concretamente, usaremos el test de rangos de Wilcoxon [17] como procedimiento estadístico no paramétrico para realizar comparaciones

entre dos algoritmos y, para comparaciones múltiples, usamos el test de Iman-Davenport [17] para detectar diferencias estadísticas entre un grupo de resultados y el test post-hoc de Holm [9] para encontrar qué algoritmos rechazan la hipótesis de igualdad con respecto a un método de control seleccionado.

## 4 Análisis Experimental

Los resultados de los experimentos para el algoritmo FH-GBML se muestran en la Tabla 2 donde, por columnas, podemos observar los resultados de entrenamiento y test para el algoritmo FH-GBML básico y para FH-GBML con los enfoques de aprendizaje por parejas, tanto para el criterio de clasificación de no dominancia (notada con el sufijo ND), y para la regla de decisión basada en una estrategia de voto (notada con el sufijo EV).

Dividimos nuestro análisis experimental en dos partes:

- Primero, queremos determinar si la propuesta del multi-clasificador mejora el rendimiento del SCBRD lingüístico que maneja todas las clases independientemente.
- A continuación, nuestro objetivo es analizar el comportamiento de nuestra propuesta para el proceso de decisión de la salida basada en un criterio de no dominancia contra una estrategia de voto.

### 4.1 Análisis de la utilidad de la propuesta de aprendizaje por parejas para sistemas de clasificación basados en reglas difusas lingüísticas

Para llevar a cabo el primero estudio, mostramos el ranking de los enfoques FH-GBML asignando las posiciones de los algoritmos en función del rendimiento para cada conjunto de datos y obteniendo a continuación el valor medio. La Figura 1 muestra el ranking medio calculado para las tres diferentes alternativas: el enfoque básico y las dos técnicas multi-clasificador. Podemos observar que tanto los métodos FH-GBML-ND como FH-GBML-EV son la mejor opción, puesto que el FH-GBML básico obtiene el peor ranking con un valor mucho mayor que los anteriores.

A continuación, realizamos un test de Iman-Davenport (Tabla 3) para detectar diferencias significativas entre los resultados de estas propuestas. El “p-value” asociado es cercano a cero, lo que implica que hay diferencias significativas entre los resultados y así, debemos aplicar un test post-hoc, en este caso el test de Holm. Queremos comparar el algoritmo FH-GBML básico contra ambos enfoques de aprendizaje por parejas.

Tabla 2: Resultados medios en acierto para el algoritmo FH-GBML con el enfoque básico y con la versión de multi-clasificación

Conjunto	#Cl.	FH-GBML		FH-GBML-ND		FH-GBML-EV	
		Entr.	Test	Entr.	Test	Entr.	Test
bal	3	85.64 ± 2.49	82.24 ± 2.85	87.04 ± 1.97	84.96 ± 2.55	86.84 ± 2.12	84.96 ± 2.55
iri	3	99.33 ± 0.70	93.33 ± 4.08	99.50 ± 0.46	94.67 ± 2.98	98.17 ± 1.09	94.00 ± 2.79
new	3	96.74 ± 0.66	91.16 ± 3.03	99.53 ± 0.49	95.81 ± 1.95	93.72 ± 1.26	94.42 ± 1.27
win	3	97.61 ± 0.63	92.70 ± 4.21	100.0 ± 0.00	96.08 ± 3.75	98.31 ± 1.46	94.40 ± 5.23
veh	4	62.44 ± 1.07	58.15 ± 3.47	74.53 ± 1.09	66.67 ± 4.46	73.82 ± 0.77	66.08 ± 4.57
cle	5	63.05 ± 0.73	50.84 ± 6.14	75.42 ± 0.76	56.56 ± 1.83	73.06 ± 0.69	57.92 ± 1.96
pag	5	95.67 ± 0.51	94.53 ± 0.91	98.17 ± 0.49	95.62 ± 1.20	96.49 ± 0.94	95.62 ± 1.77
shu	5	95.52 ± 0.67	95.22 ± 1.43	98.45 ± 0.60	97.70 ± 0.83	95.53 ± 1.46	95.40 ± 2.16
gla	6	70.44 ± 0.37	60.29 ± 6.30	82.48 ± 0.72	68.25 ± 7.34	73.83 ± 1.48	63.11 ± 6.28
seg	7	79.82 ± 1.19	78.70 ± 2.27	94.96 ± 0.35	93.29 ± 1.73	79.26 ± 3.45	78.23 ± 2.55
eco	8	79.76 ± 1.44	76.19 ± 5.04	92.34 ± 1.27	81.55 ± 4.63	84.67 ± 0.97	76.80 ± 6.48
pen	10	71.93 ± 1.19	69.82 ± 1.84	96.34 ± 0.67	91.09 ± 1.14	96.82 ± 0.35	92.09 ± 2.05
yea	10	53.84 ± 1.26	51.22 ± 4.54	65.40 ± 0.66	58.90 ± 1.44	61.12 ± 0.61	57.14 ± 1.94
let	26	18.19 ± 1.64	16.35 ± 1.58	86.30 ± 0.31	71.55 ± 2.18	46.81 ± 2.11	44.45 ± 6.27
Mean	X	76.43 ± 1.04	72.20 ± 3.41	89.32 ± 0.70	<b>82.34 ± 2.72</b>	82.75 ± 1.34	78.19 ± 3.42

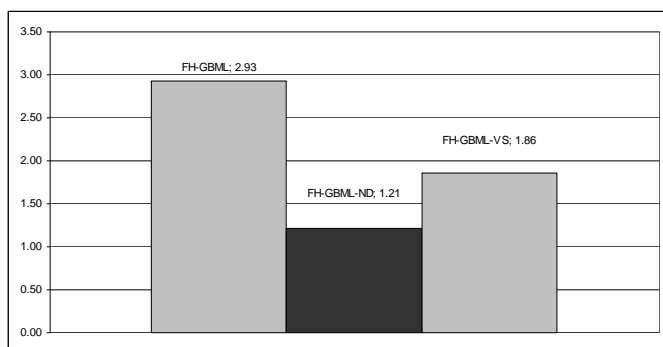


Figura 1: Ranking en media para las diferentes versiones FH-GBML: esquema básico, multi-clasificador con criterio de no dominancia y multi-clasificador con estrategia de voto

Tabla 3: Resultados del test de Iman-Davenport para la comparativa del algoritmo FH-GBML con los enfoques básico y multi-clasificador. El estadístico se calcula con 2 y 26 grados de libertad para  $\alpha = 0.05$  y 14 conjuntos de datos

Iman-Davenport	Valor en $F_F$	p-value
<b>39.0</b>	3.3690	1.49012E-8

El resultado de este test se muestra en la Tabla 4, en la que podemos observar que el algoritmo FH-GBML es superado en rendimiento por la versión de multi-clasificador, tanto para FH-GBML-ND como para FH-GBML-EV. Podemos concluir por tanto que el uso de la propuesta de aprendizaje por parejas proporciona un beneficio en el rendimiento para SCBRDs lingüísticas.

Tabla 4: Test de Holm para los enfoques FH-GBML. FH-GBML es el método de control

$i$	algoritmo	$z$	$p$	$\alpha/i$	Hipótesis ( $\alpha = 0.05$ )
2	FH-GBML-ND	4.536	5.745E-6	0.025	Rech. FH-GBML-ND
1	FH-GBML-EV	2.835	0.005	0.05	Rech. FH-GBML-EV

## 4.2 Análisis del proceso de decisión basado en no dominancia contra una estrategia de voto

El siguiente objetivo de este análisis empírico es estudiar las dos alternativas seleccionadas para el proceso de decisión entre todas las predicciones de los clasificadores difusos. Con esta meta, hemos llevado a cabo un test de Wilcoxon, mostrado en la Tabla 5, en el que comparamos el criterio de no dominancia (ND) contra la estrategia de voto (EV).

Tabla 5: Test de Wilcoxon para comparar la propuesta de no dominancia contra la regla de decisión que usa una estrategia de voto para el algoritmo FH-GBML.  $R^+$  corresponde a la suma de los rangos para el criterio de no dominancia y  $R^-$  a la regla de decisión que usa una estrategia de voto

Comparativa	$R^+$	$R^-$	p-value	Hipótesis ( $\alpha = 0.05$ )
FH-GBML-ND vs. FH-GBML-EV	92.5	12.5	0.012	Rechazado para FH-GBML-ND

Debemos destacar que el criterio de no dominancia (con un rendimiento promedio de 82.34) sobrepasa estadísticamente a la alternativa de la regla de decisión basada en la estrategia de voto (con un rendimiento promedio de 78.19), lo que supone una clara ventaja del enfoque propuesto para el proceso de decisión en sistema multi-clasificadores lingüísticos.

Adems, de acuerdo a los resultados ordenados por número de clases en la Tabla 2, podemos observar que ambas metodologías son muy similares en rendimiento para conjuntos de datos con un número bajo de clases (hasta 6), pero entonces la calidad de los resultados para la estrategia de voto decrece en contraste con el criterio de no dominancia.

## 5 Conclusiones

En esta contribución, hemos aplicado una metodología de aprendizaje de parejas para construir un sistema multi-clasificador lingüístico orientado a discriminar entre parejas de clases y para obtener una mejor frontera de decisión en problemas multi-clase.

Para agregar la salida de cada clasificador simple, hemos usado una relación de preferencia difusa, traduciendo el problema de clasificación en un problema de toma de decisión. Para obtener la clase final de salida, hemos propuesto el uso de una regla de decisión basada en el criterio de no dominancia maximal, y hemos contrastado el comportamiento de este modelo con una estrategia de voto.

El estudio experimental mostró dos conclusiones principales: primero, la aplicación de una propuesta de aprendizaje por parejas mejora claramente el rendimiento del método de SCBRD lingüísticas. Segundo, hemos determinado que la mejor alternativa para la clasificación final es el uso de nuestro criterio propuesto de no dominancia en este contexto, especialmente cuando el número de clases del problema es alto.

## Agradecimientos

Este trabajo de investigación ha sido posible gracias a la subvención del Ministerio Español de Ciencia y Tecnología bajo los Proyectos TIN2008-06681-C06-01 y TIN2007-65981.

## Referencias

- [1] J. Alcalá-Fdez, L. Sánchez, S. García, M. del Jesus, S. Ventura, J. Garrell, J. Otero, C. Romero, J. Bacardit, V. Rivas, J. Fernández, and F. Herrera. KEEL: A software tool to assess evolutionary algorithms to data mining problems. *Soft Computing*, 13(3):307–318, 2009.
- [2] A. Asuncion and D. Newman. UCI machine learning repository, 2007.
- [3] J. Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30, 2006.
- [4] A. Fernández, M. Calderón, E. Barrenechea, H. Bustince, and F. Herrera. Enhancing fuzzy rule based systems in multi-classification using pairwise coupling with preference relations. In *EUROFUSE09 Workshop on Preference Modelling and Decision Analysis*, pages 39–46, 2009.
- [5] J. Fürnkranz. Round robin classification. *Journal of Machine Learning Research*, 2:721–747, 2002.
- [6] S. García, A. Fernández, J. Luengo, and F. Herrera. A study of statistical techniques and performance measures for genetics-based machine learning: Accuracy and interpretability. *Soft Computing*, 13(10):959–977, 2009.
- [7] S. García and F. Herrera. An extension on “statistical comparisons of classifiers over multiple data sets” for all pairwise comparisons. *Journal of Machine Learning Research*, 9:2677–2694, 2008.
- [8] T. Hastie and R. Tibshirani. Classification by pairwise coupling. *The Annals of Statistics*, 26(2):451–471, 1998.
- [9] S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6:65–70, 1979.
- [10] J. C. Hühn and E. Hüllermeier. FR3: A fuzzy rule learner for inducing reliable classifiers. *IEEE Transactions on Fuzzy Systems*, 17(1):138–149, 2009.
- [11] E. Hüllermeier and K. Brinker. Learning valued preference structures for solving classification problems. *Fuzzy Sets and Systems*, 159(18):2337–2352, 2008.
- [12] H. Ishibuchi, T. Nakashima, and M. Nii. *Classification and modeling with linguistic information granules: Advanced approaches to linguistic Data Mining*. Springer-Verlag, 2004.
- [13] H. Ishibuchi and T. Yamamoto. Rule weight specification in fuzzy rule-based classification systems. *IEEE Transactions on Fuzzy Systems*, 13:428–435, 2005.
- [14] H. Ishibuchi, T. Yamamoto, and T. Nakashima. Hybridization of fuzzy GBML approaches for pattern classification problems. *IEEE Transactions on System, Man and Cybernetics B*, 35(2):359–365, 2005.
- [15] J. Luengo and F. Herrera. Domains of competence of fuzzy rule based classification systems with data complexity measures: A case of study using a fuzzy hybrid genetic based machine learning method. *Fuzzy Sets and Systems*, in press, doi:10.1016/j.fss.2009.04.001, 2009.
- [16] S. Orlovsky. Decision-making with a fuzzy preference relation. *Fuzzy Sets and Systems*, 1:155–167, 1978.
- [17] D. Sheskin. *Handbook of parametric and non-parametric statistical procedures*. Chapman & Hall/CRC, second edition, 2006.