

Rough Sets Divisibles Basados en Clustering Jerárquico

Rocío Martínez López, Miguel A. Sanz Bobi

Computer Science Department, Engineering School-ICAI,
Pontificia Comillas University
Alberto Aguilera, 25
Madrid, 28015
{rmartinez,masanz}@upcomillas.es

Resumen

Desde que la teoría de rough sets fue propuesta por Pawlak en 1982, se ha aplicado en la extracción de conocimiento de bases de datos de todo tipo. Sin embargo, se han detectado ciertos inconvenientes en casos complejos y de gran inconsistencia, como su falta de flexibilidad y su excesiva dependencia respecto a la discretización inicial de los atributos. Para superarlos aquí se propone una nueva técnica híbrida, denominada “rough sets divisibles”, que combina rough sets de precisión variable con clustering jerárquico y perceptrones. Así se pretende analizar las clases de equivalencia que mediante rough sets no generan ninguna regla con objeto de ver si mediante alguna división se puede obtener alguna nueva regla cierta. Además, se incorpora el concepto de centro de clase de equivalencia, útil en la asignación de decisiones a ejemplos que no caen en ninguna región positiva y para la división de clases de equivalencia. Las comparaciones del método de rough sets clásico y el de los “rough sets divisibles” ante varios casos ejemplo muestran que este último ofrece mejores resultados en la mayoría de los tests realizados.

Palabras clave: Rough sets, clustering jerárquico, extracción automática de conocimiento.

1. Introducción

La teoría de rough sets [1-3] fue desarrollada por Zdzislaw Pawlak a comienzos de los 80. Es una potente herramienta matemática encuadrada en el área del aprendizaje automático inductivo que ofrece recursos matemáticos para descubrir patrones ocultos en los datos. Además, es capaz de asignar incertidumbre al conocimiento extraído e identificar dependencias totales o parciales (relaciones causa-efecto) en bases de datos eliminando datos redundantes y pudiéndose emplear en problemas de clasificación.

Desde que se propuso la teoría de rough sets ésta ha sido estudiada por multitud de investigadores y ha experimentado considerables progresos [4]. Se ha

aplicado en casos de medicina [5], ingeniería [6], finanzas [7], etc.

Asimismo, se han creado hibridaciones entre rough sets y otros métodos matemáticos que mejoran la calidad de las reglas de decisión inducidas por el método de rough sets [8]. También, la reducción de los datos basándose en rough sets puede ser muy útil en el preprocesamiento de los datos de entrada a redes neuronales. Así, se han desarrollado métodos híbridos entre rough sets y redes neuronales, que conducen a nuevos modelos de neuronas [9]. Igualmente, la optimización basada en programas evolutivos puede generar de manera eficiente construcciones de rough sets como reductos, patrones de datos y reglas de decisión [10]. Por otra parte, la hibridación de rough sets con métodos clásicos como el análisis de componentes

principales o métodos bayesianos genera clasificadores de mejor calidad.

En este artículo se presenta una propuesta de mejora de algunos aspectos de la técnica de extracción de conocimiento basada en rough sets de precisión variable (VPRS) [11] mediante su combinación con otras técnicas de inteligencia artificial. Se crea así una nueva técnica híbrida denominada “rough sets divisibles” para cubrir las carencias de la original y que pretende aprovechar las posibilidades que proporcionan cada una de las técnicas que la forman: el método de clustering jerárquico DIANA [12] y el perceptrón. Además, para la necesaria discretización inicial de los datos se empleará el método Chi2 Modificado [13] con algunas variaciones.

La organización del artículo es como sigue. En primer lugar, se expondrán los conceptos teóricos de los “rough sets divisibles”. Una vez definidos los pasos de la nueva técnica, se mostrará un ejemplo de su utilización y se probará su efectividad en la extracción de conocimiento en diversos conjuntos de datos reales para realizar una comparativa con los resultados obtenidos por otras técnicas.

2. Estrategia del Algoritmo de los Rough Sets Divisibles

La extracción de conocimiento basada en rough sets adolece de ciertos inconvenientes fundamentalmente cuando trabaja con conjuntos de datos grandes y muy inconsistentes. Entre ellos se encuentran la falta de flexibilidad y la excesiva dependencia respecto a los intervalos elegidos en la discretización de los atributos. Aquí se propone un nuevo algoritmo que trata de superarlos realizando dos mejoras: una relativa a las clases de equivalencia obtenidas con el método de rough sets y otra posterior realizada a partir de nuevos ejemplos reservados del conjunto de datos.

El objetivo será perfeccionar el aprendizaje fundamentalmente en las clases de equivalencia pertenecientes a la región límite de las que no se ha podido obtener ninguna regla cierta con la aplicación del modelo VPRS. Para lograrlo, se incorpora el concepto de centro de una clase de equivalencia. Estos centros, extraídos tras un proceso de clustering de los ejemplos que las

componen, serán útiles en la partición de las clases de equivalencia.

Así, habrá dos posibles divisiones de las clases de equivalencia, una realizada a partir de los centros obtenidos en el clustering y otra que trabaja con los nuevos ejemplos de actualización del conocimiento y que asimismo tiene en cuenta dichos centros. Se habrán creado “subclases de equivalencia” que estarán definidas por los valores discretizados de los atributos de condición y por nuevos atributos generados mediante ecuaciones matemáticas en las que participan los atributos sin discretizar. Estos nuevos subconjuntos podrán generar tanto nuevas reglas positivas como inciertas.

3. Algoritmo de los Rough Sets Divisibles

Para llevar a cabo el método de los rough sets divisibles, se ha desarrollado un algoritmo que se compone de los siguientes pasos:

- Paso 1: Creación de la tabla de decisión: se distribuyen los ejemplos del conjunto de datos a tratar en una tabla.
- Paso 2: Extracción inicial del conocimiento: se emplea el método VPRS y se refinan los resultados a partir de un clustering de las clases de equivalencia.
- Paso 3: Actualización del conocimiento: mediante hiperplanos se separan los ejemplos que son más cercanos a otras clases de equivalencia que a la propia.
- Paso 4: Test: se prueban las reglas finales obtenidas con nuevos ejemplos.

3.1. Paso 1: Creación de la Tabla de Decisión

El objetivo es expresar el conjunto de datos del que se quiere extraer conocimiento de forma que pueda ser tratado en los siguientes pasos del algoritmo.

Para ello, se seleccionan el atributo de decisión, que clasifica los ejemplos, y los atributos de condición, que son los factores que se considera que pueden realizar esta clasificación. El objetivo final será determinar el valor del atributo de decisión a partir de la información proporcionada por los atributos de condición, es decir, obtener las reglas de conocimiento subyacentes que gobiernan la relación entre estos atributos.

Al igual que en el método de rough sets, los ejemplos se proporcionan al algoritmo en una tabla de decisión en la que se distribuyen por filas los ejemplos disponibles para el entrenamiento y en la que cada columna se corresponde con uno de los atributos considerados. Cada celda de la tabla indica el valor que tiene un elemento en uno de estos atributos. Este valor vendrá expresado tanto en forma discreta (necesaria para el método de los rough sets) como numérica y normalizada (necesaria para los procesos de clustering y actualización de conocimiento del algoritmo).

3.2. Paso 2: Extracción Inicial del Conocimiento

El objetivo es el descubrimiento de las reglas encubiertas en el conjunto de datos. Se empleará el método VPRS y a continuación se realizará un proceso de agrupamiento o clustering mediante el método DIANA con el fin de obtener nuevo conocimiento con un mayor número de reglas ciertas. Así, este paso lo componen dos fases:

- En la primera se aplica el método de rough sets con un nivel de error admisible de clasificación, tal y como propone el modelo VPRS.
- En la segunda se realizará el clustering en cada una de las clases de equivalencia resultantes no incluidas en la región positiva y que no han generado reglas ciertas.

3.2.1. Fase 1

Se comentan brevemente los conceptos matemáticos de esta fase que se corresponden con el modelo de rough sets de precisión variable [11].

Sea $A = \{c_1, \dots, c_n\}$ el conjunto de atributos de condición y $B = \{d\}$ el conjunto de atributos de decisión. El conjunto A tiene asociada la relación de equivalencia:

$$\tilde{A} = \{(x, y) \in U \times U : f(x, c_a) = f(y, c_a) \forall c_a \in A\}$$

donde U es el conjunto de ejemplos de entrenamiento y $f(x, c_a)$ es el valor discretizado que toma el ejemplo x para el atributo c_a . La relación de equivalencia \tilde{A} induce la partición $A^* = \{X_1, \dots, X_n\}$ sobre U que divide los ejemplos en clases de equivalencia. De la misma forma, se obtiene la partición $B^* = \{Y_1, \dots, Y_m\}$ formada por bloques e inducida por la relación de equivalencia asociada a la decisión.

Teniendo en cuenta un nivel de error admisible de clasificación β , tal que $0 \leq \beta \leq 0.5$, se definen las regiones β -positiva y β -límite de la partición B^* a partir de las β -aproximaciones superior e inferior de $Y_j \in B^*$ en el espacio de aproximación $S=(U, \tilde{A})$:

$$POSS_{\beta}(B^*) = \bigcup_{Y_j \in B^*} \underline{S}_{\beta}(Y_j) = \bigcup_{Y_j \in B^*} \left(\bigcup_{X_i \subseteq X} X_i \right)$$

$$BND S_{\beta}(B^*) = \bigcup_{Y_j \in B^*} \left(\overline{S}_{\beta}(Y_j) - \underline{S}_{\beta}(Y_j) \right) =$$

$$= \bigcup_{Y_j \in B^*} \left(\bigcup_{\beta < c(X_i, X) < 1-\beta} X_i \right)$$

Cada clase de equivalencia encuadrada en la región positiva generará una regla cierta entre los atributos de condición y la decisión. Por el contrario, las clases de equivalencia incluidas en la región límite darán lugar a reglas inciertas con un grado de incertidumbre expresado mediante un factor de confianza. Por lo tanto, se obtendrá como resultado de esta fase:

1. Reglas ciertas de la forma $X_i \rightarrow Y_j$, donde la condición es una clase de equivalencia y la conclusión es un bloque de la partición inducida por el atributo de decisión.
2. Reglas inciertas de la forma $X_i \rightarrow Y_j$ con un factor de confianza calculado:

$$\alpha_{Y_j}(X_i) = \frac{|X_i \cap Y_j|}{|X_i|}$$

donde $|A|$ indica la cardinalidad de un conjunto finito A .

Puesto que se ha permitido un nivel de error admisible de clasificación $\beta > 0$, las reglas ciertas tendrán un error de clasificación comprendido en el intervalo $[0, \beta]$. Asimismo, ninguna regla incierta tendrá $\alpha_{Y_j}(X_i) > 1 - \beta$.

3.2.2. Fase 2

El modelo de rough sets finaliza en la fase anterior la extracción de conocimiento de los datos disponibles. Sin embargo, cuando se trabaja con tablas de decisión muy inconsistentes, este método puede considerarse demasiado rígido, a pesar de la ligera flexibilidad proporcionada por el nivel de error admitido β . Para solucionarlo, se propone añadir una segunda fase que permita la división en “grupos” de las clases de equivalencia pertenecientes a la región límite de la partición inducida por el atributo de decisión. El objetivo será la obtención de nuevas reglas ciertas a partir de las clases de equivalencia que únicamente han generado reglas inciertas.

Al finalizar este proceso, cada ejemplo dispondrá de un atributo más, el centro del grupo de su clase de equivalencia al que es más cercano. Además, los centros de cada grupo serán útiles para el siguiente paso de actualización de conocimiento.

En la fase anterior se asignaron las clases de equivalencia X_i de la partición A^* a las regiones positiva o límite de la partición B^* . Ahora, se tratarán todas y cada una de estas clases según la región a la que pertenezcan de la siguiente forma:

Caso 1: La clase X_i está incluida en la región positiva ($POSS\beta(B^)$)*

Se calcula un centro M media de todos los ejemplos que componen la clase de equivalencia. Se considerará que la clase está compuesta por un único grupo.

Caso 2: La clase X_i está incluida en la región límite ($BND\beta(B^)$)*

Se calcularán uno o varios centros para la clase de equivalencia tratada mediante clustering con DIANA. Estos centros se emplearán para dividir a X_i en grupos realizando una inducción ficticia de una partición.

Sea $P = \{p_1, p_2, \dots, p_c\}$ el conjunto de centros obtenido, siendo cada p_k un punto definido por sus valores en los atributos de condición. Se dividirá a la clase X_i en tantos grupos como centros tenga el conjunto P . Sea $G = \{G_1, G_2, \dots, G_c\}$ el conjunto de grupos que se van a formar a partir de X_i , conteniendo cada G_k un subconjunto de ejemplos de la clase tratada X_i y teniendo además asociado el punto centro p_k .

Para distribuir todos los casos ejemplo pertenecientes a X_i en los distintos grupos de G , se empleará la medida de la distancia euclídea. Así, un ejemplo $x \in X_i$ formará parte del grupo G_k que tenga asociado el centro p_k más cercano a x . Por lo tanto, el índice k del grupo al que debe pertenecer un ejemplo x , se calculará:

$$k(x) = \arg_k \min \|x - p_k\|$$

Una vez que se han repartido todos los ejemplos del conjunto X_i en los diversos grupos G_k , se cumple que G es un conjunto de subconjuntos no vacíos de X_i y además se verifica que $G_1 \cup G_2 \cup \dots \cup G_c = X_i$ y $G_i \cap G_j = \{\}$ para $i \neq j$. Por lo tanto, G posee las

características necesarias para poder considerarse una partición de la clase X_i , y en consecuencia, se podrá tratar a cada G_k como un bloque de la partición G . Los elementos que pertenezcan a uno de estos bloques están caracterizados por tener los mismos valores discretizados en todos los atributos de condición y por coincidir en el atributo ficticio añadido (ser más cercanos al centro asociado al grupo que a los centros del resto de bloques).

A continuación, se comprobará para cada uno de los bloques de la partición G si puede incorporarse a la región positiva o límite de la partición B^* . Así, los bloques de G que pertenezcan a la región positiva, generarán reglas ciertas, y los que estén encuadrados en la región límite, generarán reglas inciertas con cierto grado de incertidumbre expresado mediante un factor de confianza, tal que:

$$\alpha_{y_j}(G_k) = \frac{|G_k \cap Y_j|}{|G_k|}$$

Sin embargo, la forma de expresar la regla es más compleja porque se debe añadir el concepto de que el ejemplo debe estar más próximo al centro correspondiente que al resto. En consecuencia, una regla obtenida a partir del grupo G_k será: "si el atributo c_1 es xxx, ..., y el atributo c_n es zzz, y además, el ejemplo es más cercano al centro p_k que al resto de centros de la clase de equivalencia, entonces el atributo d es ddd".

Finalmente, si se ha generado alguna nueva regla cierta, se ha cumplido el objetivo principal de esta fase, obtener algún conocimiento cierto a partir de unos ejemplos de los que no se podía extraer una conclusión clara de a qué decisión pertenecían. Era necesario el atributo añadido (centro más cercano) para obtener estas reglas ciertas.

Cálculo de centros mediante DIANA.

El método DIANA es un algoritmo de clustering jerárquico que comienza con un único cluster que contiene todos los objetos de la clase de equivalencia X_i y que va subdividiendo sucesivamente en clusters más pequeños de manera que se vaya formando un dendrograma.

Para agilizar el proceso de agrupamiento jerárquico y no perder tiempo en una excesiva ramificación del dendrograma, se impone una nueva condición de parada al método, la selección de un número final máximo de clusters n menor al número de ejemplos que componen X_i .

Tras ejecutar el método, se tienen los n clusters y se realiza una selección de los más representativos. Para ello, se calcula la representatividad de un cluster final J :

$$REP_J = \frac{|J|}{\max(d(i, j))}$$

siendo $\max(d(i, j))$ la máxima distancia entre dos de los ejemplos del cluster.

A continuación, se seleccionan el cluster final que tenga la mayor representatividad, siendo ésta \max_REP_j , y también los clusters que cumplan que: $REP_j > \frac{1}{2} * \max_REP_j$. Finalmente, se calcula para cada uno de estos clusters el centro media de todos los ejemplos que lo componen, y estos centros pasarán a formar parte del conjunto de centros final P .

3.3. Paso 3: Actualización del conocimiento

El objetivo de este paso es obtener más conocimiento no generado en el paso anterior. Para ello se analizarán, a partir de nuevos ejemplos de actualización, las zonas de las clases de equivalencia que estén cercanas a grupos de las vecinas.

Sea $Q = \{x_1, x_2, \dots, x_n\}$ el conjunto de datos reservado para la actualización de conocimiento de los que no se conoce la decisión. Para cada $x_i \in Q$ se realizará el siguiente proceso:

En primer lugar, se comprueba a qué clase de equivalencia $X_i \in A^*$ pertenecería el ejemplo x_i según sus valores discretizados de los atributos. Además, siendo $G = \{G_1, G_2, \dots, G_c\}$ el conjunto de grupos de la clase de equivalencia creado en la fase 2 de la extracción de conocimiento, se mirará a qué grupo pertenece según el centro más próximo, siendo p_k el centro asociado al grupo G_k y $k(x_i)$ el índice del grupo elegido.

Si el ejemplo no pertenece a un grupo o “subclase de equivalencia” encuadrado en la región positiva, se calcularán sus distancias euclídeas a los centros de los grupos de las clases de equivalencia vecinas. Si el centro más cercano está más próximo que el del propio grupo al que pertenece el ejemplo, se contabiliza el caso.

Los ejemplos originales de entrenamiento de un grupo se dividirán en el momento en el que se contabilice un porcentaje determinado de ejemplos de actualización que estén más cercanos a otro grupo que al propio. Este porcentaje, calculado con

respecto al número de ejemplos que componen el grupo, se determina utilizando el método de prueba y error siendo muy dependiente del conjunto de datos estudiado. La división se realizará mediante la ecuación del hiperplano equidistante al centro del propio conjunto y al del conjunto al que los ejemplos de actualización son más cercanos.

Este hiperplano es equivalente al definido por una neurona de tipo perceptrón. De esta forma, en lo sucesivo, en los grupos divididos de este modo, para comprobar a qué subconjunto de los dos creados pertenece un ejemplo nuevo, se presentará al perceptrón asociado, y según su salida se asignará a un subconjunto o al otro.

Finalmente, se comprueba si cada uno de los subconjuntos generados en la división pertenece a la región positiva o a la región límite y se calcula su centro medio por si se requiriera posteriormente su división.

3.4. Paso 4: Test

Sea $T = \{x_1, x_2, \dots, x_n\}$ el conjunto de datos reservado para test. Para cada ejemplo de test $x_i \in T$ se comprueba a qué clase de equivalencia $X_i \in A^*$ pertenecería el ejemplo según sus valores discretizados de los atributos. Además, se mirará a qué grupo pertenecería según el centro que esté más próximo de entre los calculados en la fase 2 de extracción inicial de conocimiento. Igualmente, si este grupo ha sido dividido por uno o más hiperplanos, se presentará el ejemplo al perceptrón correspondiente.

Una vez conocido el grupo al que pertenece el ejemplo, si éste está encuadrado en la región positiva, se le asignará directamente la decisión correspondiente. Por el contrario, si pertenece a la región límite, se le asignará la decisión con mayor factor de certeza. En el caso de que el ejemplo de test no pertenezca a ninguna de las clases de equivalencia constituidas en la extracción de conocimiento, se le asignará la decisión del grupo perteneciente a la región positiva cuyo centro sea el más cercano.

4. Ejemplo

A continuación, en la figura 1, se muestra un conjunto de datos de ejemplo caracterizado por dos atributos de condición (c_1 y c_2) y un atributo de

decisión con dos posibles valores (positivo y negativo). Sobre este conjunto se aplicará el algoritmo descrito.

La tabla de decisión obtenida al ejecutar el paso 1 del algoritmo se muestra en la Tabla 1.

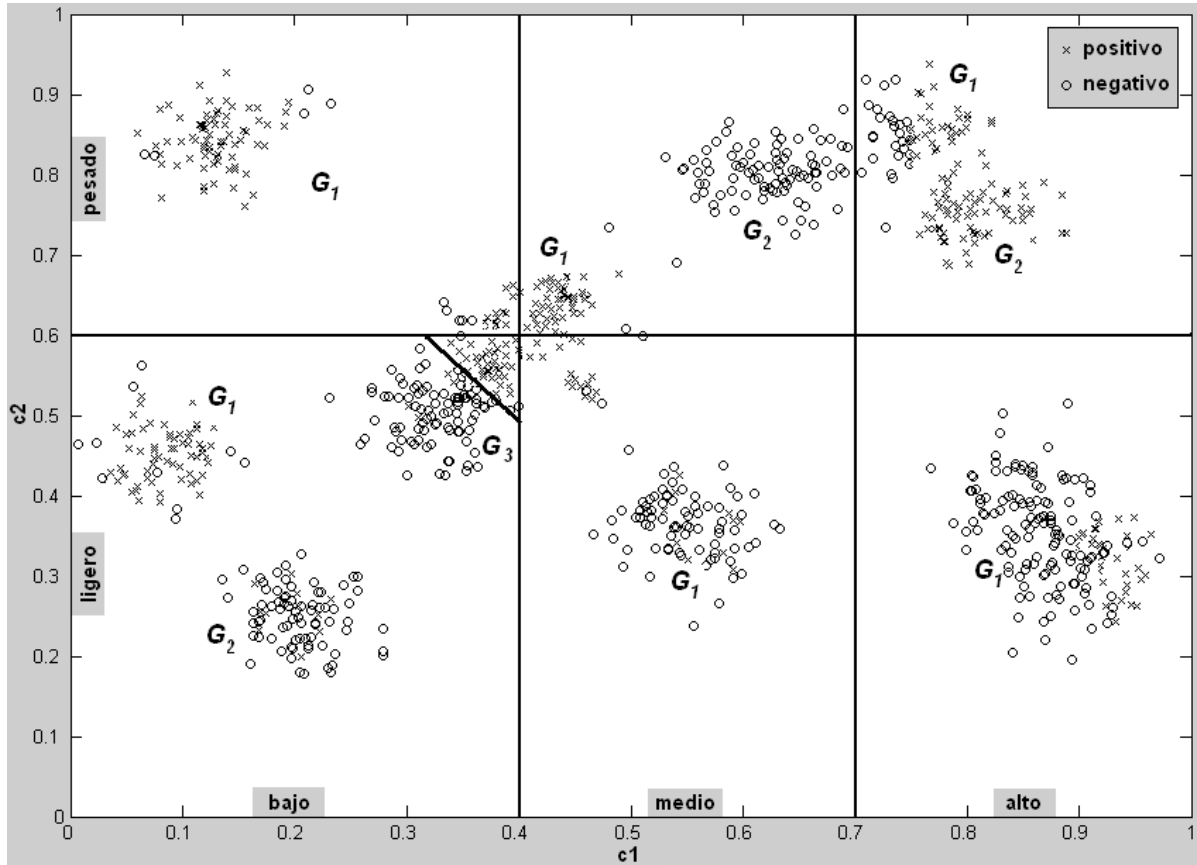


Figura 1. Representación gráfica del conjunto de datos de ejemplo

	Atributo c_1	Atributo c_2	Decisión d
Ejemplo 1	0.45 (medio)	0.63 (pesado)	positivo
Ejemplo 2	0.22 (bajo)	0.34 (ligero)	negativo
...	... (...)	... (...)	...

Tabla 1. Tabla de decisión del conjunto de datos de ejemplo

En el paso 2 se generan las particiones $A^* = \{X_1, X_2, X_3, X_4, X_5, X_6\}$, donde X_1 contiene los datos con c_1 bajo y c_2 ligero, etc., y $B^* = \{Y_1, Y_2\}$, donde Y_1 son los datos con decisión negativa y Y_2 son los que la tienen positiva. Algunas de las reglas son:

- Si c_1 es alto y c_2 es ligero, entonces d es negativo (regla completamente cierta)
- Si c_1 es bajo y c_2 es pesado, entonces d es positivo (regla completamente cierta)
- Si c_1 es bajo y c_2 es ligero, entonces d es

negativo con $\alpha_{Y_1}(X_1) = |X_1 \cap Y_1| / |X_1| = 0.35$.

- Si c_1 es bajo y c_2 es ligero, entonces d es positivo con $\alpha_{Y_2}(X_1) = |X_1 \cap Y_2| / |X_1| = 0.65$.

Para la clase de equivalencia con c_1 bajo y c_2 ligero (X_1), se puede observar que se obtienen tres clusters representativos (G_1, G_2, G_3) cada uno con un centro (p_1, p_2, p_3) . G_1 y G_2 se incorporarán a la región positiva de B^* , generando dos reglas ciertas, mientras que G_3 formará parte de la región límite. Las nuevas reglas generadas serán:

- Si c_1 es bajo, c_2 es ligero, y además el ejemplo es más cercano al centro p_1 que a p_2 y p_3 , entonces d es positivo (regla completamente cierta).
- Si c_1 es bajo, c_2 es ligero, y además el ejemplo es más cercano al centro p_2 que a p_1 y p_3 , entonces d es negativo (regla completamente cierta).
- Si c_1 es bajo, c_2 es ligero, y además el ejemplo es más cercano al centro p_3 que a p_1 y p_2 , entonces d es negativo con un factor de confianza = 0.70.
- Si c_1 es bajo, c_2 es ligero, y además el ejemplo es más cercano al centro p_3 que a p_1 y p_2 , entonces d es positivo con un factor de confianza = 0.30.

En el paso 3 puede darse el caso de que haya ejemplos de actualización que pertenezcan al grupo G_3 de la clase de equivalencia X_1 , pero que sean más cercanos al centro del grupo G_1 de la clase de equivalencia con c_1 medio y c_2 pesado. Cuando el porcentaje de estos ejemplos de actualización sea significativo, se dividirá el grupo G_3 con la ecuación del hiperplano equidistante entre el centro del grupo G_3 y el del grupo G_1 . Así, G_3 queda dividido en dos subconjuntos, cada uno de los cuales generará una regla cierta puesto que la mayoría de sus ejemplos son de una única clase.

5. Resultados Experimentales

En este punto se realizará una comparativa entre los resultados proporcionados en la extracción de conocimiento realizada con VPRS y con el nuevo método de rough sets divisibles. Para ello, se trabajará con conjuntos de datos reales muy distintos del repositorio UCI de bases de datos para aprendizaje automático [14]. Estos conjuntos son ampliamente conocidos y utilizados en esta área de conocimiento, por lo que no se profundizará en su descripción: "Iris" es el más simple, "BUPA" y "Diabetes" son también conjuntos sencillos aunque no tan separables linealmente y "Glass" y "Abalone" son los menos lineales y más complejos. En el método VPRS cada conjunto se divide en el conjunto de entrenamiento (2/3) y el de test (1/3). En los rough sets divisibles, 2/3 del primero forman el conjunto de entrenamiento inicial (paso 2) y el resto es el de actualización (paso 3). En cada conjunto, los parámetros se eligen por prueba y

error para ser los más adecuados, mostrándose de forma orientativa en la Tabla 2 los parámetros que han proporcionado mejores resultados.

Tanto para el análisis mediante rough sets como con los rough sets divisibles, es necesario realizar un primer paso de discretización de las variables continuas. El método empleado ha sido el Chi2 Modificado [13] con una variación, se permitirá rebajar el nivel de consistencia L_c inicial con el propósito de obtener clases de equivalencia más representativas que tengan un mayor número de ejemplos.

Como se observa en la Tabla 2, los pasos que forman la técnica de los "rough sets divisibles" proporcionan un incremento progresivo en la tasa de aciertos obtenida, para finalmente superar a los rough sets de precisión variable en cuatro de los cinco casos tratados. Esto se debe principalmente a que:

1. A los ejemplos de test que no pertenecen a ninguna clase de equivalencia de las formadas en el entrenamiento se les adjudica la decisión del grupo incluido en la región positiva con el centro más cercano. Esto se puede realizar a partir de la fase 2 de la extracción inicial de conocimiento, cuando se tienen los centros calculados.
2. Un mayor número de ejemplos cae en reglas ciertas (muchos en las generadas tras las divisiones de las clases de equivalencia), con lo que se puede asignar con mayor seguridad a más ejemplos una decisión considerada como cierta.

En la Tabla 3 se realiza una comparativa de los resultados obtenidos para los conjuntos de datos estudiados cuando se emplean otros métodos de extracción de conocimiento que siguen estrategias distintas a la adoptada por los rough sets.

Por una parte se tienen dos tipos de árboles de decisión (ID3 [16], la única que al igual que los rough sets trabaja con los atributos discretizados, y CART [17]) y por otra, dos tipos de redes neuronales (perceptrón multicapa [18] y redes de base radial [19]). En la mayoría de los casos estas técnicas superan los resultados obtenidos por los rough sets de precisión variable, mientras que se observa que ofrecen resultados similares a los alcanzados por los rough sets divisibles.

Conjuntos	Características			Parámetros			Tasa de aciertos			
	Ejemplos	Atributos	Clases	L_c final	n clusters máximo	% ejem. actualiz.	RS Div. P2-F1	RS Div. P2-F2	RS Div. P3	VPRS
Iris	150	4	3	0.75	3	10%	92.00%	94.00%	94.00%	92.00%
BUPA	345	6	2	0.58	4	10%	48.69%	65.22%	66.96%	54.78%
Abalone	4177	4	9	0.53	8	15%	54.67%	54.60%	54.74%	55.68%
Glass	214	7	6	0.40	3	20%	56.34%	67.61%	69.01%	57.75%
Diabetes	768	7	2	0.65	4	20%	55.47%	67.58%	67.97%	59.77%

Tabla 2. Conjuntos de datos estudiados y tasas de aciertos

Conjuntos	Tasa de aciertos			
	ID3	CART	Perceptrón multicapa	RBFN
Iris	94.00%	94.00%	96.00%	96.00%
BUPA	60.87%	64.35%	70.43%	60.00%
Abalone	55.03%	54.24%	56.03%	55.75%
Glass	67.61%	70.42%	54.93%	54.93%
Diabetes	74.22%	69.53%	73.05%	65.63%

Tabla 3. Tasas de aciertos con otros métodos de extracción de conocimiento

Por una parte se tienen dos tipos de árboles de decisión (ID3 [16], la única que al igual que los rough sets trabaja con los atributos discretizados, y CART [17]) y por otra, dos tipos de redes neuronales (perceptrón multicapa [18] y redes de base radial [19]). En la mayoría de los casos estas técnicas superan los resultados obtenidos por los rough sets de precisión variable, mientras que se observa que ofrecen resultados similares a los alcanzados por los rough sets divisibles.

6. Conclusiones

En este artículo se ha propuesto una nueva técnica híbrida, los "rough sets divisibles", fruto de la combinación de los rough sets de precisión variable con neuronas de tipo perceptrón y la técnica de clustering jerárquico DIANA. Además se ha empleado para la discretización de los datos el método Chi2 modificado, realizando algunas innovaciones en su utilización con el fin de obtener reglas más significativas.

Con esta técnica se pretende la obtención de reglas ciertas a partir de las reglas inciertas aportadas por

el método de rough sets. De esta forma, al presentar nuevos ejemplos a la base de conocimiento generada, aumentará el número de éstos a los que se puede asignar una decisión considerada como cierta. Además, se ha permitido la asignación de una clase a los ejemplos que caen en reglas inciertas o que no caen en ninguna regla porque en el entrenamiento ningún ejemplo era similar a ellos.

La técnica exige la elección de varios parámetros como son el nivel de consistencia alcanzado en la discretización, el número final máximo de clusters de cada clase de equivalencia o el porcentaje mínimo de ejemplos de actualización requerido para proceder a la división de una clase de equivalencia. Los valores asignados a estos parámetros deberán ser adecuados al caso tratado, teniendo en cuenta que una mala selección de éstos puede causar la generación de reglas menos significativas.

Finalmente, como futura ampliación a la técnica descrita se propone la incorporación de redes neuronales RBFN en sustitución del perceptrón empleado en la fase de actualización de conocimiento. Estas redes permitirían una división no lineal de los ejemplos que compongan una clase de equivalencia perteneciente a la región límite, con

la consiguiente mejora en los resultados obtenidos en conjuntos de datos más complejos.

Referencias

1. Pawlak, Z., "Rough Sets". *International Journal of Computer and Information Sciences*, vol. 11, no. 5, pp. 341-356. Oct. 1982.
2. Pawlak, Z., *Rough Sets: Theoretical Aspects of Reasoning About Data*. Dordrecht, Netherlands: Kluwer Academic, 1991.
3. Pawlak, Z., Grzymala-Busse, J., Slowinski, R., Ziarko, W. "Rough Sets". *Communications of the ACM*, vol. 38, no. 11, pp. 89-95. Nov. 1995.
4. Slowinski, R., *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory*. Kluwer Academic Publishers, Boston, 1992.
5. Tsumoto, S., Tanaka, H., "Automated Discovery of Medical Expert System Rules from Clinical Databases Based on Rough Sets". *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD' 96)*, AAAI Press, Menlo Park, pp. 63-69. 1996.
6. Nowicki, R., Slowinski, R., y Stefanowski, J., "Evaluation of Vibroacoustic Diagnostic Symptoms by means of the Rough Sets Theory". *Computers in Industry*, vol. 20. pp. 141-152. 1992.
7. Ziarko, W., Golan, R. y Edwards, D. "An Application of Datalogic/R Knowledge Discovery Tool to Identify Strong Predictive Rules in Stock Market Data". *Proceedings of AAAI Workshop on Knowledge Discovery in Databases*, Washington, DC, 1993. pp. 93-101.
8. Browne, C., Düntsch, I., Gediga, G., *IRIS Revisited, A Comparison of Discriminant and Enhanced Rough Set Data Analysis*. En: [15], pp. 345-368. 1998.
9. Szczuka, M., *Rough sets and artificial neural networks*. En: [15], pp. 451-471. 1998.
10. Wróblewski, J., *Genetic algorithms in decomposition and classification problems*. En: [15], pp. 472-492. 1998.
11. Ziarko, W., "Variable Precision Rough Set Model. *Journal of Computer and System Sciences*", vol. 46, no. 1, pp. 39-59. Feb. 2003.
12. Kaufman, L., Rousseeuw, P.J. *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: John Wiley & Sons, pp. 253-279, 1990.
13. Shen, L., Tay, F.E.H., "A discretization method for Rough Sets Theory". *Intelligent Data Analysis*, vol. 5, pp. 431-438. 2001.
14. C.J. Merz and P.M. Murphy. *UCI repository of machine learning databases*. <http://www.ics.uci.edu/~mllearn/MLRepository.html>. Irvine, CA: University of California, Department of Information and Computer Science, 1996.
15. Polkowski, L., Skowron, A. (Eds.), *Rough Sets in Knowledge Discovery*, Heidelberg: Physica Verlag, vol. 2, 1998.
16. Quinlan, J.R. "Induction of decision trees," *Machine Learning*, vol. 1, pp. 81-106, 1986.
17. Breiman, L., Friedman, J.H., Olshen, R.A. y Stone, C.J. *Classification and regression trees*. Monterey, CA: Wadsworth International, 1984.
18. Rumelhart, D.E., Hinton G.E. y Williams, R.J. "Learning representations by back-propagating errors". *Nature*, vol. 323, pp. 533-536, 1986.
19. Moody, J. y Darken, C.J. "Fast learning in networks of locally-tuned processing units," *Neural Computation*, vol. 1, pp. 281-294, 1989.