

Minería de datos con conjuntos aproximados para clasificación de imágenes satelitales*¹

[Data mining with rough sets for classification of satellite images]

JUAN OLEGARIO MONROY VÁSQUEZ²

RECIBO: 20.08.2011 - AJUSTE: 20.09.2011 - AJUSTE: 15.11.2011 - APROBACIÓN: 10.12.2011

Resumen

Este artículo trata del estado de arte frente al uso de Rough Set en la clasificación de imágenes satelitales, ésta técnica hace parte de un conjunto de algoritmos que se agrupan dentro de lo que se denomina minería de datos. Hasta ahora Rough set se ha aplicado preferencialmente en el descubrimiento de insolvencias en datos obtenidos de manera experimental durante un lapso de tiempo específico, característica que ha llevado a que se implemente con éxito en empresas que requieren tomar decisiones basadas en los análisis de cifras de producción en periodos de tiempo determinados; de este conjunto de algoritmos de minería de datos, muchos se han implementado en la clasificación de imágenes satelitales buscando determinar o bosquejar elementos presentes en la superficie terrestre de acuerdo a su comportamiento frente

* Modelo para citación de este artículo de revisión:
MONROY VÁSQUEZ, Juan Olegario (2011). Minería de datos con conjuntos aproximados para clasificación de imágenes satelitales. En: Ventana Informática. No. 25 (jul. – dic., 2011). Manizales (Colombia): Facultad de Ciencias e Ingeniería, Universidad de Manizales. p. 129-158. ISSN: 0123-9678

- 1 Artículo proveniente del proyecto titulado *Análisis de la aplicación de la metodología Rough Set en la clasificación de imágenes satelitales*, iniciado en agosto de 2009, e inscrito en el grupo de investigación Gicoge, para optar al título de Magíster en Ciencias de la Información y las Comunicaciones, de la Universidad Distrital Francisco José de Caldas, bajo la dirección del PhD. José Nelson Pérez Castillo.
- 2 Ingeniero Electrónico; Especialista en Multimedia Educativa; MSc(c) en Ciencias de la Información y las Comunicaciones.
Miembro del grupo GICOGE de la Universidad Distrital; Director del grupo de investigación PRODIG de la Universidad Nacional Abierta y a Distancia; Miembro del grupo GLUNAD de la Universidad Nacional Abierta y a Distancia. Docente Auxiliar Universidad Nacional Abierta y a Distancia. Sogamoso (Colombia). Correo electrónico: juan.monroy@unad.edu.co

a la radianza electromagnética y así distinguir patrones dentro de las imágenes.

Palabras Clave: *Conjuntos Aproximados, Imágenes Satelitales, indiscernible*

Abstract

This article is about art state in front of The Rough Set use in the satellite images classification, this technique takes part of a set of techniques and algorithms which are grouped in the data mining. Till now Rough set has been applied preferentially in the discovery of information insolvencies obtained in an experimental way during a specific time, this characteristic has let it to be implemented successfully in companies which need to make decisions based on the production numbers analysis in certain periods of time; many of these data mining algorithms set have been implemented in the classification of the satellite images seeking to determine or to draw the elements on the earth surface according to its behaviour towards the electromagnetic radiance and in this way to distinguish patterns inside these images.

Keywords: *Rough Set, Satellite Images, indiscernible*

Introducción

Actualmente hay un gran conjunto de algoritmos que dan respuesta a la necesidad de descubrimiento de patrones en un conjunto de datos como los recogidos por imágenes satelitales, estos algoritmos están clasificados dentro de las técnicas de minería de datos (descubrimiento semiautomático de patrones, asociaciones, cambios, anomalías y estructuras estadísticamente significantes y eventos en los datos, según Grossman et al. (1998); estos algoritmos se agrupan en dos ramas: aquellos que hacen uso de ecuaciones y los basados en la lógica, aunque algunos pueden hacer uso tanto de ecuaciones como de la lógica.

Este documento en la primera parte trata de los fundamento de la técnica de Conjuntos aproximados, destacando como su uso ha permitido que se constituya en una excelente herramienta para caracterizar información que presenta alto grado de incertidumbre, situación cotidiana en datos geográficos, administrativos y en general en la información recolectada experimentalmente; posteriormente se muestran los aspectos más relevantes a tener en cuenta en el

momento de tomar datos de imágenes satelitales para descubrir patrones y como la técnica de conjuntos aproximados pueden contribuir en el proceso.

1. Conjuntos aproximados (Rough Set)

La base de la teoría de los conjuntos aproximados está en la suposición de que cada objeto del universo de discurso tiene rasgos característicos, los cuales son presentados por información (conocimiento, datos) acerca del objeto, de acuerdo con Pawlak (2002). Los objetos que tienen las mismas características son indiscernibles. La teoría ofrece herramientas matemáticas para descubrir patrones escondidos en los datos, identifica dependencias parciales o totales, es decir relaciones causa – efecto, en bases de datos, elimina redundancia en los datos, da aproximaciones a valores nulos o inválidos, datos perdidos, datos dinámicos etc., considera Pawlak (1991).

La teoría de la Técnica de *Rough Set* surgió en la década de los 80 propuesta por Zdzislaw Pawlak, desarrollada originalmente como una herramienta para el análisis y clasificación de datos, y hoy en día se ha aplicado con éxito en diversas tareas, tales como la selección, extracción, definición de reglas de síntesis y clasificación, expresan Komorowski et al. (1998), con lo que se abrió una nueva dirección en el desarrollo de teorías sobre la información incompleta.

Esta metodología se clasifican dentro de las técnicas de minería de datos (descubrimiento semiautomático de patrones, asociaciones, estructuras estadísticamente significantes y eventos en los datos, afirman Grossman et al. (1998), a su vez, se pueden dividir en aquellas que hacen uso de ecuaciones, tales como la estadística o las redes neuronales, y las que se basan en la lógica como los árboles de decisión y las reglas, sin embargo, algunas pueden hacer uso tanto de ecuaciones como de la lógica. La técnica *Rough Set* está dentro de las metodologías basadas en la lógica, según afirman Adriaans et al. (1996), Bonilla y Esteban (2003) y Adriaans (2002).

1.1 Conceptos

La técnica de Conjuntos Aproximados se puede abordar desde el significado de aproximado (*Roughness*), que Pawlak (1991) define como el opuesto al de exactitud, donde la aproximación representa el grado del completo conocimiento que se tenga de la clasificación de un conjunto dado, de manera que si la aproximación es 0, el conocimiento de la clasificación puede definirse como el de un conjunto exacto.

La filosofía de *Rough Set* se basa en suponer que a cada objeto del universo que se considere, se puede asociar alguna información (datos, conocimiento), situación que en algunos casos conduce a objetos no discernibles a la vista de los datos disponibles, ésta relación de no diferenciación genera la base matemática para esta teoría, donde, la información imprecisa, es la causa de la no diferenciación de los objetos y evita en consecuencia su asignación precisa a un conjunto. *Rough Set* es un conjunto de objetos que, en general, no pueden ser caracterizados de manera precisa en términos de la información disponible, expresan Arco y Bello (2006).

El objetivo de la teoría *Rough Set* consiste en descubrir dependencias entre atributos a partir de la información disponible y organizada en tablas, esto, con el fin de reducir la cantidad de elementos que forman el conjunto, eliminando aquellos que no son esenciales para caracterizar el conocimiento, de manera que se logren subconjuntos de atributos que aseguren la misma calidad de clasificación, que si se dispusiera de todos ellos, a estos subconjuntos se les denomina reductos, según plantean Pawlak (1982) y Düntsch y Gediga (2000).

En las filas de la tabla 1 se indican los objetos (acciones, alternativas, candidatos, pacientes, empresas, etc.), en las columnas se ubican los atributos, a partir de Düntsch y Gediga (2000). Las entradas en la tabla son los valores del atributo; es decir, la entrada en la columna q y en fila p tiene el valor $f(p, q)$, por tanto, para cada par (objeto, atributo) se conoce un valor denominado descriptor. Cada fila de la tabla contiene descriptores que representan información correspondiente a un objeto. La relación de no-diferenciación ocurriría si dados dos objetos sus descriptores tomaran el mismo valor.

En relación con los atributos, Düntsch y Gediga (2000), los clasifica en: atributos de condición (criterios, pruebas, síntomas, etc.) y en atributos de decisión (decisiones, clasificaciones, taxonomías, etc.).

Tabla 1. Representación de una tabla de información (Modificado de Düntsch y Gediga, 2000)

Objetos	Atributos		
	1	2	N
1	Valor (1, 1)	Valor (1, 2)	Valor (1, n)
2	Valor (2, 1)	Valor (2, 2)	Valor (2, n)
M	Valor (m, 1)	Valor (m, 2)	Valor (m, n)

1.2 Reducción y dependencia de atributos

Como se mencionó anteriormente, una de las funciones más importantes del trabajo con *Rough Set* es el descubrimiento de las dependencias entre atributos al analizar una tabla de información, esto, porque permite reducir el conjunto al eliminar aquellos atributos que sean redundantes, llegando así al reducto o conjunto mínimo. En una tabla de información puede haber más de un reducto y su intersección da el denominado núcleo que, para Zopounidis y Dimitras (1998) y Caballero, et al. (2007a), es la colección de los atributos más relevantes en la tabla y que no pueden ser eliminados sin que disminuya la calidad de la aproximación en la clasificación.

Una forma de reducción de los datos se observa en la cantidad de objetos del universo a considerar y la otra en la reducción de la cantidad de atributos a tener en cuenta. La esencia es mantener o considerar solo aquellos atributos que preservan la relación de inseparabilidad; los atributos rechazados son redundantes y su eliminación no afecta la calidad de la inferencia. El cálculo de los reductos de un sistema de decisión es un proceso computacionalmente costoso; esto ha llevado al desarrollo de algunas variantes para el concepto de reducto, presentados por Zhong, Dong y Ohsuga (2001), Kohavi y Frasca (1994), Yao, Wong y Butz (1999) y Caballero et al. (2007b).

Al final un conjunto aproximado puede ser reemplazado o representado por un par de conjuntos precisos, llamados aproximación por encima y por debajo, que según Yun y Yuanbin (2004), la aproximación por encima corresponde a todos los objetos que con seguridad pertenecen al conjunto y la aproximación por debajo contiene los elementos que posiblemente pertenecen al conjunto.

1.3 Reglas de decisión

Con *Rough Set* se obtienen las reglas de decisión de un sistema de información; una regla de decisión puede expresarse como una sentencia lógica que relaciona la descripción de condiciones y las clases de decisión, su estructura se muestra a continuación:

SI <se cumplen condiciones> ENTONCES <el objeto pertenece a una clase de decisión dada>

Las reglas generadas pueden ser determinísticas o no determinísticas. Para Segovia et al. (s.f.), por determinística (consistente, precisa, exacta) se entiende si $C \rightarrow D$ y en otro caso, es no determinística (inconsistente, aproximada) que sería cuando las condiciones pueden conducir a varias posibles decisiones.

1.4 Ejemplo de aplicación

A continuación se muestra como se aplica la técnica de *Rough Set* a un conjunto de datos que pretenden definir si el desempeño de un deportista es bueno medio ó malo, según su trabajo en tres pruebas de su estado físico, una prueba de actividad moderada, una prueba aeróbica, y una anaeróbica. Los resultados son medidos y clasificados como bueno, medio y bajo y codificados como 1,0 y -1 respectivamente. El proceso se puede resumir en cinco pasos: - Definir la Tabla de información, - Definir la tabla de decisión (Tabla 3), - Establecer la discernibilidad de los elementos del conjunto a aproximar, - Descubrir las dependencias y obtener las reducciones, y - Obtener las reglas de decisión.

1.4.1 Definir la tabla de información. Se ubican en una Tabla 2 los datos que caracterizan a cada uno de los objetos que hacen parte del estudio.

Tabla 2. Ejemplo Tabla de información (Construido a partir de Hor y Crossley, 2005)

Deportista	Pruebas		
	Aeróbica	Anaeróbica	Moderada
D1	1	-1	1
D2	0	1	1
D3	-1	-1	-1
D4	0	-1	-1
D5	0	-1	-1
D6	1	1	-1

1.4.2 Definir la tabla de decisión. Se adiciona a la tabla de información en atributo de decisión que permite en conjunto caracterizar los objetos (Tabla 3).

Tabla 3. Ejemplo Tabla de decisión (Construido a partir de Hor y Crossley, 2005)

Deportista	Atributos de condición			Atributos de decisión
	Aeróbica	Anaeróbica	Moderada	Estado
D1	1	-1	1	1
D2	0	1	1	-1
D3	-1	-1	-1	-1
D4	0	-1	-1	1
D5	0	-1	-1	-1
D6	1	1	-1	1

Aquí, simplemente, se han definido cuales se consideran los atributos de decisión y cuales los de condición, en este caso el atributo de deci-

sión es aquel que caracteriza al deportista como bueno o malo, según su desempeño en las actividades de prueba, en donde se categorizan como malo ó medio para conformar así los atributos de condición.

La tabla de decisión de manera formal esta conformada por una cuadrupla así:

$$S = \langle U, Q, V, f \rangle$$

Donde:

$U = \{x_1, x_2, \dots, x_n\}$ conjunto finito de objetos, en este caso los seis deportistas

$$U = \{D1, D2, D3, D4, D5, D6\}$$

$C = \{q(1), q(2), \dots, q(m)\}$ conjunto finito de atributos de condición

$$C = \{\text{Aeróbica}, \text{Anaeróbica}, \text{Moderada}\}$$

$D = \{q(m+1), q(m+2), \dots, q(m+n)\}$ conjunto finito de atributos de decisión

$$D = \{\text{Estado}\}$$

$Q = \{q_1, q_2, \dots, q_m\}$ conjunto finito de atributos

$$Q = \{\text{Aeróbica}, \text{Anaeróbica}, \text{Moderada}, \text{Estado}\}$$

V_q es el dominio de cada atributo q , V es la unión de todos los dominios:

$V = \{-1, 0, 1\}$ que corresponde a -1 =malo, 0 = medio y 1 bueno el desempeño

$f: U \times Q \rightarrow V$ función tal que $f(x, q) \in V_q$, para $x \in U$, llamada función de información.

$f(\text{Objeto}, \text{atributo})$

$$f(D1, \text{Aeróbica})=1, f(D1, \text{Anaeróbica})=-1, f(D1, \text{Moderada})=1, \dots$$

1.4.3 Establecer la discernibilidad de los elementos del conjunto aproximado. Una vez definida la tabla de decisión, es necesario establecer las relaciones de *indiscernibilidad*, que se definen como:

Se dice que x está relacionado con y , se escribe: $xI_P y$, si para todo atributo $q \in P$, donde P es un subconjunto de atributos de V , se cumple:

$$f(x_1, q) = f(x_2, q)$$

Entonces para un conjunto P de atributos, $P=\{\text{aeróbica, anaeróbica, moderada}\}$ se tiene los elementos discernibles e indiscernibles:

$$I_p(D1)=\{D1\}$$

$$I_p(D2)=\{D2\}$$

$$I_p(D3)=\{D3\}$$

$$I_p(D4)=I_p(D5)=\{D4,D5\}$$

$$I_p(D6)=\{D6\}$$

$$I_p=\{\{D1\},\{D2\},\{D3\},\{D4,D5\},\{D6\}\}$$

En el caso anterior los elementos indiscernibles son el D4 y D5; ahora las aproximaciones con un conjunto X1 de deportistas cuyo estado es bueno=1 son: la inferior $\underline{PX1}(x)$ y la superior $\overline{PX1}(x)$ dadas como:

$$\underline{PX1}(x) = \{x \in U : I_p(x) \subseteq X1\}$$

$$\underline{PX1}(x) = \{D1, D6\}$$

$$\overline{PX1}(x) = \bigcup_{x \in X} I_p(x)$$

$$\overline{PX1}(x) = \{D1, D4, D5, D6\}$$

Y la Frontera

$$B_p X1(x) = \overline{PX1}(x) - \underline{PX1}(x)$$

$$B_p X1(x) = \{D4, D5\}$$

Ahora las aproximaciones con un conjunto X2 de deportistas cuyo estado es malo=-1 son: la inferior $\underline{PX2}(x)$ y la superior $\overline{PX2}(x)$ dadas como:

$$\underline{PX2}(x) = \{x \in U : I_p(x) \subseteq X2\}$$

$$\underline{PX2}(x) = \{D2, D3\}$$

$$\overline{PX2}(x) = \bigcup_{x \in X} I_p(x)$$

$$\overline{PX2}(x) = \{D2, D3, D4, D5\}$$

Y la Frontera

$$B_p X2(x) = \overline{PX2}(x) - \underline{PX2}(x)$$

$$B_p X2(x) = \{D4, D5\}$$

1.4.4 Descubrir las dependencias y obtener las reducciones. Es necesario establecer ahora las dependencias y obtener las reducciones,

para ello se parte de la calidad de las aproximaciones encontradas, para el presente caso se tiene el conjunto aproximado $X=\{X1,X2\}$, cuya calidad de aproximación se da por:

$$\gamma_p(X) = \frac{|\underline{P}(X1)| + |\underline{P}(X2)|}{|U|}$$

Donde:

$||$ Indica la *cardinalidad* del conjunto que contiene

$\gamma_p(X1)$ Representa la fracción de objetos de U correctamente Clasificados

Entonces:

$$\gamma_p(X) = \frac{|\underline{P}(X1) + \underline{P}(X2)|}{|U|} = \frac{2+2}{6} = 0.667$$

Para obtener las dependencias y reducciones es necesario tener la calidad de aproximación para cada atributo, que se muestran en la tabla 4, donde los conjuntos de atributos son:

$P1=\{Aeróbica\}$

$P2=\{Anaeróbica\}$

$P3=\{Moderada\}$

$P4=\{Aeróbica, Anaeróbica\}$

$P5=\{Aeróbica, Moderada\}$

$P6=\{Anaeróbica, Moderada\}$

$P=\{Aeróbica, Anaeróbica, Moderada\}$

γ Calidad de la Aproximación.

Tabla 4. Ejemplo calidad de las aproximaciones para cada atributo
(Construido a partir de Hor y Crossley, 2005)

	P1	P2	P3	P4	P5	P6	P
γ	0.5	0.0	0.0	0.667	0.667	0.5	0.667

De la tabla 4 se concluye que los reductos de P son P4 y P5, que son los conjuntos de atributos:

$P4=\{Aeróbica, Anaeróbica\}$

$P5=\{Aeróbica, Moderada\}$

Son reductos porque la calidad de aproximación de cada uno de ellos es igual a la obtenida con P que utiliza todos los atributos $P=\{\text{Aeróbica, Anaeróbica, Moderada}\}$, de aquí se obtiene el núcleo de P, es decir, aquel atributo que no puede dejar de considerarse, el más significativo:

$$\text{NucX}(P) = P4 \cap P5 = \text{Aeróbica}$$

Significa que se puede emplear como conjunto de atributos P4 o P5 ya que se obtiene la misma calidad de aproximación que usando P.

1.4.5 Obtener las reglas de decisión. Las reglas de decisión, como se mencionó anteriormente, son el tipo Si. ENTONCES, las cuales se construyen a partir de los atributos reductos P4 ó P5 o, considerando todos los atributos, P así:

Para $P = \{\text{Aeróbica, Anaeróbica, Moderada}\}$, que es el conjunto formado por los tres atributos tienen seis reglas de decisión así:

Si $f(x, P1)=1$ y $f(x, P2)=-1$ y $f(x, P3)=1$ ENTONCES
 $f(x, \text{Estado}) = 1 = \text{Bueno}$

Si $f(x, P1)=0$ y $f(x, P2)=1$ y $f(x, P3)=1$ ENTONCES
 $f(x, \text{Estado}) = -1 = \text{Malo}$

Si $f(x, P1)=-1$ y $f(x, P2)=-1$ y $f(x, P3)=-1$ ENTONCES
 $f(x, \text{Estado}) = -1 = \text{Malo}$

Si $f(x, P1)=0$ y $f(x, P2)=-1$ y $f(x, P3)=-1$ ENTONCES
 $f(x, \text{Estado}) = 1 = \text{Bueno}$

Si $f(x, P1)=0$ y $f(x, P2)=-1$ y $f(x, P3)=-1$ ENTONCES
 $f(x, \text{Estado}) = -1 = \text{Malo}$

Si $f(x, P1)=1$ y $f(x, P2)=1$ y $f(x, P3)=-1$ ENTONCES
 $f(x, \text{Estado}) = 1 = \text{Bueno}$

De las anteriores reglas se concluye que cuatro de ellas son determinísticas (1, 2, 3, 6) y dos no lo son (4, 5); dado que estas últimas reglas de decisión no permiten determinar el valor que debe asignársele a un deportista según los tres atributos considerados en P, hace que los elementos con estas características hagan parte de la frontera definida anteriormente.

$$B_p X1(x) = \overline{P}X1(x) - \underline{P}X1(x)$$

$$B_p X1(x) = \{D4, D5\}$$

Y por tanto, es aconsejable ampliar el rango de atributo que hagan posible discernir a los elementos de hacen parte de dicha frontera.

Para $P4 = \{\text{Aeróbica}, \text{Anaeróbica}\}$, que es el conjunto formado por los dos atributos reductos, se tienen seis reglas de decisión así:

SI $f(x, P1)=1$ y $f(x, P2)=-1$ ENTONCES
 $f(x, \text{Estado}) = 1 = \text{Bueno}$

SI $f(x, P1)=0$ y $f(x, P2)=1$ ENTONCES
 $f(x, \text{Estado}) = -1 = \text{Malo}$

SI $f(x, P1)=-1$ y $f(x, P2)=-1$ ENTONCES
 $f(x, \text{Estado}) = -1 = \text{Malo}$

SI $f(x, P1)=0$ y $f(x, P2)=-1$ ENTONCES
 $f(x, \text{Estado}) = 1 = \text{Bueno}$

SI $f(x, P1)=0$ y $f(x, P2)=-1$ ENTONCES
 $f(x, \text{Estado}) = -1 = \text{Malo}$

SI $f(x, P1)=1$ y $f(x, P2)=1$ ENTONCES
 $f(x, \text{Estado}) = 1 = \text{Bueno}$

De las anteriores reglas se concluye que cuatro de ellas son determinísticas (1, 2, 3, 6) y dos no lo son (4, 5); dado que estas últimas reglas de decisión e igual que en el caso anterior no permiten discernir los elementos, se sugiere entonces ampliar el rango de atributos.

1.5 Algunas aplicaciones de la técnica de Rough Set

Rough Set se ha aplicado preferencialmente en el descubrimiento de insolvencias en conjuntos de datos obtenidos por lo general, de manera experimental durante un lapso de tiempo determinado, por ello se ve como se ha implementado con éxito en el análisis de datos recogidos por empresas en determinados periodos de producción y a partir de los cuales se deben tomar decisiones de tipo gerencial, principalmente en el análisis de la insolvencia empresarial, según lo expresan Segovia et al. (s.f.), Segovia et al. (2005) y Díaz, Fernández y Segovia (2004).

Además de estas aplicaciones de la cuales existen amplia y variada evidencia en la web, puede encontrarse otras áreas de aplicación como:

- Descubrimiento de conocimiento útil de datos geográficos múltiples donde el tamaño y la complejidad del conjunto de datos desafían la comprensión humana, corroborado como un estudio de caso que se concentran en conocimientos sobre la distribución de animales y sobre el peligro de avalancha (Colin, 1998).
- Modelamiento del sistema de un motor ac y técnicas para el diagnóstico de fallas (Bonaldi et al., 2002).
- Propuesta de un modelo de computación granular basado en conjuntos aproximados (Ying, Peiyao y Dongsheng, s.f.).
- Como herramienta para inferir acerca de la correspondencia de imágenes de rostros con diferencias en la iluminación (Yong et al., s.f.).
- La teoría de conjuntos aproximados con tolerancia provee una forma de agrupar los resultados de la búsqueda en la web (Ngo y Nguyen (2005) y Wakaki, Itakura y Tamura (2004)).
- En aspectos meteorológicos se ha implementado para clasificar los diferentes tipos de tormentas (Peters et al., 2003).
- Los detectores de metales utilizados en la actualidad por los equipos dedicados a la descontaminación de la minería, de manera que es necesario confirmar los datos sensados a partir de el descubrimiento de las reglas en las bases de datos de las minas, aquí *Rough Set* se convierte en una herramienta aplicable (Choudharil y Nandi, 2005)
- Minería de datos sobre texto es otra de las aplicaciones en las que se ha trabajado, donde además se ha desarrollado un nuevo concepto *Rough Text* en la validación de agrupamientos, el etiquetado de agrupaciones y el resumen de multidocumentos (Arco y Bello, 2006).
- El RSSM, como un servicio basado en *Rough Set* con el objetivo de tolerar la incertidumbre en la búsqueda de servicios en la *grid*. Los resultados de la evaluación han mostrado que la RSSM algoritmo es más eficaz en el descubrimiento que otros mecanismos como UDDI y OWL (Li et al., 2006)
- En la medicina se usa *Rough set* como una herramienta de apoyo al análisis de datos médicos enmarcados dentro de los estándares de medida (Ohrn, 1999).
- Preprocesamiento de los datos, para editar conjuntos de entrenamiento en la solución de problemas de clasificación. Se propone un algoritmo para la edición de dichos conjuntos, basado en los conceptos de aproximaciones de la RST y la técnica *Bootstrap* (Cárdenas, Caballero y Bello, 2007)
- *Rough Set* también se ha empleado junto con otras técnicas para crear algoritmos híbridos, por ejemplo se ha utilizado para apoyar un

método de construcción de la estructura de una red neuronal difusa (Xian, Ji-Kai y Yan-Hong, s.f.).

- Creación de un servicio web inteligente para clasificar imágenes digitales empleando conjuntos aproximados (Caicedo y Pérez, 2010)

Por otra parte, en la Universidad de Poznan (Polonia) se desarrolló el programa *ROSE2, Rough Set Data Explorer*, según lo presentan Predki et al. (1998), y Predki y Wilk (1999).

2. Clasificación de imágenes satelitales (Raster)

El objetivo de la clasificación de imágenes satelitales es identificar los elementos presentes en la superficie terrestre, para ello se basa en información captada y discretizada por un sistema de sensores de radianza electromagnética ubicados en satélites, esta radianza se traduce en un entero entre 0 y 255 por cada pixel de la imagen satelital; con esta información se alimenta un algoritmo que define los patrones o categorías superficiales en la que entran los elementos que la componen y les asigna un falso color para conformar la imagen de la clasificación de la cobertura vegetal del terreno.

Para comprender mejor el concepto de clasificación que se enunció anteriormente, es necesario abordar algunos aspectos referentes a las imágenes satelitales, los tipos de satélites de teledetección que actualmente se utilizan para captar información de la superficie terrestre, satélite *Landsat 7*, ampliar el proceso de clasificación y, finalmente, enunciar el software de clasificación de imágenes satelitales más destacado.

2.1 Características de las imágenes satelitales

Una imagen satelital corresponde, afirma Felicísimo (2001), a una imagen digital que representa gráficamente un objeto sobre la superficie de la tierra mediante una matriz regular que recoge valores de reflectancia, los que se miden mediante sensores.

Dentro de las características a distinguir en una imagen satelital en función del objetivo del proyecto están: la potencialidad de la resolución frente al uso que se le da al resultado del procesamiento de dichas imágenes y la información genérica a extraer según la banda o rango de frecuencia de la onda electromagnética (su longitud de onda), captada como niveles digitales.

La Resolución Espacial hace referencia al área representada en cada uno de los píxeles que componen la imagen, donde puede encontrarse, según Delgado (2009):

- Resoluciones de 80 m. Con características de uso en: cartografía estructuras geológicas regionales, en evaluación de la salud vegetativa en una región relativamente extensa 1 kilómetro, en valoración de la salud vegetativa en estados y países enteros, seguimiento de eventos regionales como plagas de insectos, sequía y desertificación.
- Resolución espacial de 20-30 m. Con este tipo de resolución es posible ubicar aeropuertos, cascos urbanos, barriadas periféricas, centros comerciales, complejos deportivos, grandes fábricas, extensos bosques y explotaciones agrícolas de gran amplitud, de igual forma realizar clasificaciones generalizadas de la superficie del terreno,
- Resolución espacial de 10 m. A este nivel de detalle es factible ubicar y cartografiar edificios, predios, carreteras, límites de propiedad, campos de deporte, granjas y calles laterales; también permite diferenciar entre parcelas cultivadas y sin cultivar en función de la salud vegetativa relativa y por último facilitar las tipificaciones de la cubierta del suelo en pequeñas áreas.
- Resolución Espacial de 1 m. Esta resolución logra entre otras cosas: Identificar y cartografiar rasgos a la escala humana superiores a un metro cuadrado, tales como tapas de alcantarilla, bancos, automóviles, cocheras de autobuses, carriles de autopista, aceras, equipamiento de servicios públicos, cercas, árboles y arbustos. De igual forma otras tareas posibles de realizar a esta resolución son: identificar características de muchos de los objetos mencionados, detectar pequeñas zonas de estrés en parcelas agrícolas o arboledas, localizar y cartografiar ampliaciones de casas, carreteras, edificios, patios y pequeñas explotaciones agrícolas y ganaderas, diferencia distintos tipos de edificios.
- Resolución Espacial de 0.61 m.

«Estas imágenes muestran un excelente nivel de detalle, claridad, calidad espectral y cobertura global, siendo una gran herramienta para el análisis e interpretación de la superficie terrestre, especialmente en el ámbito militar, administración de recursos naturales, evaluación de desastres, planificación urbana, vialidad e infraestructura, telecomunicaciones, agricultura, control del medio ambiente, catastros, cartografía detallada, localización e identificación de zonas urbanas, minería, entre otros» (GeoService Perú, s.f., 3).
- Longitud de onda captada. Según la longitud de onda captada de la superficie terrestre, puede destacarse las principales aplicaciones

así: - Azul visible: Cartografía de aguas someras. Diferenciación de suelo y vegetación, - Verde visible: Diferenciación de la vegetación por su salud, - Rojo visible: Diferenciación de la vegetación por especies, - Infrarrojo cercano: Cartografía de la vegetación Cartografía del vigor/salud de la vegetación Diferenciación de la vegetación por especies, - Infrarrojo medio: Diferenciación de los tipos de rocas por composición, Detección de humedad en la vegetación y suelo, Cartografía de la estructura geológica, Trazado de límites tierra/agua.

2.2 Satélites de teledetección

Dentro del grupo de satélites dispuestos en órbita con fines de teledetección se encuentran aquellos que ofrecen resoluciones desde 61 cm hasta 80 m. A continuación se hace una breve descripción de los más conocidos, de acuerdo a lo presentado por CASI (2010):

- QuickBird. El satélite *QuickBird* fue lanzado en octubre de 2001, con una órbita heliosíncrona y una altura de 450 Km., pasando por un mismo lugar en un tiempo de 3 a 7 días, adquiere imágenes en blanco y negro con 61 cm de resolución e imágenes a color (4 bandas) con 2,44 m de resolución, cubriendo una superficie de 16,5 Km. x 16,5 Km.
- Ikonos. El satélite *Ikonos* fue lanzado con una órbita síncrona con el sol a una altura aproximada de 680 Km. Puede producir imágenes de un mismo lugar cada 3 días. Las imágenes tienen un ancho de 11,3 Km. en el nadir. *Ikonos* es programable y el instrumento de toma de imágenes orientable, lo que permite la revisita de un mismo sitio en menos de tres días, además de ofrecer imágenes en blanco y negro con una resolución de 1 m e imágenes en color (cuatro bandas) con una resolución de 4 m.
- OrbView. El satélite *OrbView-3* fue puesto en órbita en junio de 2003 y captura datos con una resolución espacial de 1 m en el pancromático y 4 m en el multispectral. El ancho de pasada es de 8 Km. y puede pasar por un mismo punto de la Tierra en menos de tres días, pudiendo tomar datos con un ángulo máximo de inclinación de 50°. Ofrece imágenes pancromáticas con una resolución de 1m y multispectrales (cuatro bandas) con una resolución de 4 m.
- Landsat – MSS / TM / ETM. El satélite *Landsat* se encuentra operativo desde julio de 1972. Situado a una altura de 705 Km., ofrece imágenes con una extensión de 185 Km. x 185 Km. cada 16 días. Los satélites *Landsat 4* y 5, proporcionan imágenes multispectrales de los sensores MSS (cuatro bandas) y TM (siete bandas) con una resolución

espacial de 80m y 30m, respectivamente. Además, el sensor ETM+ a bordo del *Landsat 7* ofrece una banda más pancromática de 15 m de resolución.

- Spot – HRVIR. Desde 1986 se han lanzado cinco satélites, proporcionando imágenes de media y alta resolución de la superficie de la Tierra. Los satélites Spot orbitan a una altura de 810 Km. proporcionando imágenes con una extensión de 60 Km. x 60 Km. en el nadir. Spot 5 posee nuevas características que suponen una serie de mejoras sobre los anteriores, proporcionando imágenes pancromáticas con una resolución de 5 m (2,5 m en supermodo) y multispectrales (5 m en supermodo) a una resolución de 10m.
- Formosat-2. De acuerdo con Fernández y Herrero (2001), es un satélite de alta resolución gestionado por el NSPO (*National Space Program Office* de Taiwán), lanzado el 20 de mayo de 2004, cuyos datos técnicos son: imágenes blanco y negro de 2 m de resolución, imágenes color de 8 m de resolución (cuatro bandas espectrales: infrarrojo cercano, rojo, verde y azul), cobertura de 24 Km. x 24 Km. por escena.

2.3 Satélite Landsat-7

El último satélite fue lanzado en abril de 1999 con un nuevo sensor denominado ETM+ (*Enhanced Thematic Mapper Plus*). Su operación es administrada por la NASA (*National Space and Space Administration*) y la producción y comercialización de imágenes depende de la USGS (*United States Geological Survey*).

Una imagen Landsat 7 ETM+ está compuesta por ocho bandas espectrales que pueden ser combinadas de distintas formas para obtener variadas composiciones de color u opciones de procesamiento. Entre las principales mejoras técnicas respecto a su antecesor, el satélite Landsat 5, se destaca la adición de una banda espectral (Banda Pancromática) con resolución de 15 m. También cuenta con mejoras en las características geométricas y radio métricas y una mayor resolución espacial de la banda térmica para 60 m, según afirma Conaf (1983, 76).

Estos avances tecnológicos permiten calificar al Landsat-7 como el satélite más interesante para la generación de imágenes con aplicaciones directas hasta una escala de 1:25.000, principalmente en áreas rurales o territorios de grandes extensiones. Las imágenes generadas por él presentan una mejor relación costo-beneficio que los datos generados por satélites de resolución media (15 a 30 metros) actualmente ofrecidos en el mercado.

A continuación se muestran algunas de las conclusiones de investigaciones que apuntaban al uso de imágenes satelitales *Landsat*:

- El satélite *Landsat* es el más empleado en aplicaciones forestales, agrícolas, hidrológicas, usos del suelo y monitorización medioambiental. Sobre todo, está ligado a estudios territoriales en los que el parámetro fundamental es el medio ambiente (Fao, 1986, 86)
- La Corporación Nacional Forestal de Chile, Conaf, menciona que el sistema *Landsat* permite obtener la mayor parte de la información de relevancia que requiere el sector forestal, en forma confiable, altamente periódica, entregando una visión sinóptica con nivel de precisión geométrica y cartográfica compatible con los requerimientos usuales para estudios regionales, a un costo por hectárea notablemente bajo y además, en las imágenes satelitales es posible identificar aquellas áreas críticas, en donde es necesario concentrar los esfuerzos de fomento a fin de establecer una cobertura vegetal protectora del suelo y generadora de biomasa con fines de desarrollo socioeconómico para la población, expresan González y Cuevas (1982, 46).
- Según FAO, citada por Bosque (1997, 451), el uso imágenes *Landsat* como instrumento para representar cartográficamente y evaluar las unidades de degradación del suelo, tiene las siguientes ventajas: - Permite delinear con exactitud límites fisiográficos claros, - Proporciona un medio para obtener información acerca de la tierra, cuando no se dispone de otras fuentes o cuando la calidad de la información disponible es baja y - Brinda la oportunidad de poder repetir anual o estacionalmente el examen multiespectral.

2.4 Proceso de clasificación de imágenes satelitales

Al proceso de clasificación de imágenes satelitales se ha definido como:

- El reconocimiento de clases o grupos cuyos miembros tengan ciertas características en común (Chuvienco, 2002).
- El proceso que convierte los n - Niveles digitales contenidos en una imagen multiespectral tomada por un satélite, en una imagen con un número de categorías de cobertura del suelo. Así, esta imagen permite crear cartografía de una zona en cuestión y generar información con las superficies ocupadas por cada tipo de ocupación de suelo (Jensen, 1995).

La clasificación a partir de imágenes satelitales debe ser, según Lillesand y Kiefer (1999):

- Exacta al lograr la mayor cercanía a la realidad.
- Reproducible dadas las mismas variables de entrada

- Robusta, es decir no sensible a pequeños cambios en las variables de entrada, aunque permita explotar enteramente la información de esas variables
- Exhaustiva al cubrir todo el territorio de estudio.
- Objetiva al no estar marcada por las decisiones del intérprete (aunque la experticia de éste es la que permite la identificación y corrección de las clasificaciones hechas).

Para realizar este proceso, considera Richards y Jia (2005), es necesario contar con más de una imagen satelital de la misma área y de diferente rango en el espectro electromagnético, a las que se aplica un algoritmo para obtener una nueva imagen cuyos niveles digitales (ND), no correspondan a la *radianza* detectada por el sensor, sino a una etiqueta que permita identificar la categoría asignada a ese píxel. Si la clasificación se basa exclusivamente en los ND se le denomina clasificación espectral y se puede apoyar en información que tienen que ver con la especialidad, temporalidad, y a su vez en información auxiliar (topografía, suelos, entre otros) para identificar claramente las cubiertas.

Así, el principio de clasificación digital no busca una definición absoluta de cada cubierta, aplicable a cualquier imagen, sino que ésta dependa de: la propia imagen, el sensor utilizado, la zona, fecha y tipos de cubierta. Para Lillesand y Kiefer (1999), la apariencia externa de la imagen clasificada puede ser similar a la satelital, pero sobre ella no se puede aplicar ciertas operaciones estadísticas, ya que generalmente no esta medida en escala cuantitativa, sino cualitativa. Estas imágenes clasificadas, por tanto, son sensibles a ser operables basándose en parámetros estadísticos para variables nominales (moda, índice, variación cualitativa).

En general, se puede manejar un esquema de la clasificación digital, propuesto por Gómez (2006), que involucre:

- Definición digital de las categorías (fase de entrenamiento). Esta basada en los valores numéricos de los píxeles, se busca obtener el rango de ND que identifiquen a cada categoría, para todas las bandas que intervienen en la clasificación.
- Agrupación de los píxeles de la imagen en una de las categorías (fase de asignación).
- Comprobación y verificación de resultados.

Considerando la forma como se obtienen las estadísticas de entrenamiento, se tienen: - método Supervisado, que Requiere conocimiento previo del terreno, a partir del cual se seleccionan las muestras para cada categoría, áreas representativas (*training fields*), y - No supervisado, que hace una búsqueda automática de grupos de valores homogéneos dentro de la imagen.

Ningún método permite dar solución inmediata a todos los problemas de la clasificación digital; el supervisado puede considerarse subjetivo y artificial, mientras que el no supervisado puede entregar resultados de difícil interpretación, no muy afines con las necesidades del usuario final, además de resultar poco claro el hecho que este método sea realmente capaz de identificar las agrupaciones naturales de la imagen.

3. Fase experimental

En esta fase, y como primera etapa, se pretende identificar cobertura vegetal en imágenes satelitales, para ello es necesario contar con unos atributos que permitan categorizar los elementos presentes en las imágenes, estos atributos se construyen a partir del concepto de índices de vegetación (NDVI).

Los NDVI son el resultado de combinaciones matemáticas entre los niveles digitales de *radianza* electromagnética almacenados en dos bandas espectrales de la misma región, hay diferentes alternativas de combinaciones de bandas espectrales, en general el cálculo se hace a partir del cociente entre los niveles de *radianza* presentes en cada banda; del conjunto amplio de NDVI se selecciona el que se denomina NDVI54 que se calcula con la ecuación:

$$\text{NDVI} = (\text{IR}-\text{R}) / (\text{IR}+\text{R})$$

Se selecciona este índice de vegetación, teniendo como fundamento a Speranza y Zerda (2005), quienes consideran que *«el índice de vegetación que aporta mayor información para la discriminación promedio de las coberturas forestales es el índice normalizado NDVI54, posteriormente y en forma decreciente según la capacidad de separabilidad, le siguen los índices NDVI47, NDVI53, NDVI52, NDVI51, y NDVI43; siendo el índice NDVI75 el que ofrece la menor separabilidad entre categorías, debido a la fuerte correlación entre bandas»*.

Las imágenes satelitales que se emplean para esta fase experimental son de tipo *Landsat*, se escogió la región colombiana y de ella a su vez se seleccionaron cuatro regiones, cuyo criterio de selección se basó en la variedad de elementos presentes en la superficie y que dentro de ella fuera evidente la presencia de cobertura vegetal, que ya ha sido categorizadas y reconocidas por una persona experta³. A continuación en la tabla 5 se relacionan los metadatos de la imagen Landsat seleccionada

3 Se contó con el apoyo del agrólogo Germán Herrera Salamanca, del Laboratorio de Fotogeología y Sensores Remotos de la Universidad Pedagógica y Tecnológica de Colombia UPTC seccional Sogamoso, quien reconoce y ha trabajado en dichas áreas.

Tabla 5. Metadatos imágenes Landsat Girardot (Construida a partir de GLCF, 2011)

Nombre de la imagen	Escena Landsat ETM+ , Path (8), Row (57)
Sistema de Coordenadas	WGS 84 UTM Zona 18N
Coordenadas	Mínimo en X : 907039.8682043
	Máximo en X : 924760.1317957
	Mínimo en Y : 960180.1125704
	Máximo en Y : 975819.8874296
Autor	Programa Landsat de la NASA
Editor	USGS (U.S. Geological Survey)

3.1 Datos a clasificar

Los datos que constituyen las imágenes satelitales están divididos en dos: los metadatos y los niveles de *radianza*; estos últimos datos para cada uno de los píxeles que conforman las imágenes y son datos de tipo entero en un rango de 0 a 255, con base en ellos se calcula los índices de vegetación. Es importante aclarar que los niveles de *radianza* son independientes del tamaño y la resolución de la imagen, por tanto, se utilizó un muestreo estratificado para escoger 25 datos, 5 por cada estrato así: de [0 a 50], [51 a 101], [102 a 152], [153 a 203] y [203 a 255]. Estos valores se seleccionaron de la banda espectral 5 y 4, ubicando primero la radianza de la banda 5 y con las mismas coordenadas se seleccionaba la que correspondía a la banda 4; esto se hizo con el aplicativo IDRISI. En la tabla 6 se muestra un resumen de los datos de la imagen de la región de Girardot.

La razón para seleccionar primero las *radianzas* en la banda espectral 5, es por las características de reflectividad que se presentan en este rango de frecuencia según el tipo de vegetación, tal como lo aclara Gates (2003), al indicar que el comportamiento teórico de la vegetación vigorosa muestra una reducida reflectividad en las bandas visibles debido al efecto absorbente de los pigmentos de clorofilas, xantofilas y carotenos, captando la radiación situada en torno a los 0,445 μm , mientras que las bandas espectrales del infrarrojo la reflectividad es mayor, es así como la banda 4, se usa para determinar el contenido en biomasa y para el mapeo de líneas de ribera, y la banda espectral 5 indica el contenido en agua en suelos y vegetación, penetra por nubes delgadas y presenta distintos tipos de vegetación en contrastes diferentes.

Tabla 6. Datos histogramas de las imágenes satelitales seleccionadas de la región Girardot (Construido a partir de GLCF, 2011)

Banda Espectral	Valor mín. radianza	Valor máx. radianza	Número de píxeles
4	0	255	307911
5	0	255	307911

3.2 Información imágenes Landsat Girardot

Con el procedimiento descrito anteriormente se construyó la tabla 7, que muestra la información de las imágenes *Landsat* de Girardot.

Tabla 7. Tabla de información Imágenes Landsat Girardot (Monroy, 2011)

Píxel	Columna	Fila	Nivel de Radianza	
			Banda Espectral 5	Banda Espectral 4
D1	59	221	4	75
D2	121	142	21	71
D3	515	255	39	0
D4	240	256	43	103
D5	396	154	50	99
D6	337	253	54	83
D7	152	99	62	79
D8	312	188	75	99
D9	87	40	83	103
D10	385	154	96	63
D11	390	26	104	107
D12	141	431	118	59
D13	487	75	121	136
D14	551	291	135	67
D15	255	90	147	47
D16	422	59	158	144
D17	440	89	164	148
D18	212	113	172	148
D19	444	114	185	196
D20	141	364	198	192
D21	52	330	206	176
D22	485	3	214	188
D23	562	300	220	204
D24	278	375	233	204
D25	157	343	249	204

Con base en la anterior información se estructura la tabla 8, que contiene los atributos de decisión, se construyó a partir del índice de vegetación NDVI¹⁵⁴; en la tabla de decisión no se incluyeron los valores de columna y fila, ya que no constituyen un atributo ni de condición ni de decisión.

Tabla 8. Tabla de decisión Imágenes Landsat Girardot (Monroy, 2011)

Píxel	Atributos de condición		Atributos de decisión
	Banda Espectral 5	Banda Espectral 4	NDVI
D1	4	75	-0.898734177
D2	21	71	-0.543478261
D3	39	0	1
D4	43	103	-0.410958904
D5	50	99	-0.32885906
D6	54	83	-0.211678832
D7	62	79	-0.120567376
D8	75	99	-0.137931034
D9	83	103	-0.107526882
D10	96	63	0.20754717
D11	104	107	-0.014218009
D12	118	59	0.333333333
D13	121	136	-0.058365759
D14	135	67	0.336633663
D15	147	47	0.515463918
D16	158	144	0.046357616
D17	164	148	0.051282051
D18	172	148	0.075
D19	185	196	-0.028871391
D20	198	192	0.015384615
D21	206	176	0.078534031
D22	214	188	0.064676617
D23	220	204	0.037735849
D24	233	204	0.066361556
D25	249	204	0.099337748

Los niveles de NDVI son valores entre -1 y 1, en esta tabla de decisión se calculó el *NDVI Aprox*, a partir de la aproximación a un decimal del NDVI, de manera que esta columna se convierta en el atributo de decisión. Se aproximó a 1 porque así se logran manejar 20 clases de índices de vegetación, que son suficientes para identificar la cobertura vegetal, según se comprobó en el aplicativo IDRISI manejando 16 clases.

La tabla de decisión de manera formal está conformada por una cuadrupla así:

$$S = \langle U, Q, V, f \rangle$$

Donde:

$$U = \{D1, D2, D3, D4, D5, D6, \dots, D20, D21, D22, D23, D24, D25\}$$

$$C = \{\text{Nivel de Radianza Banda Espectral 5, Nivel de radianza Banda Espectral 4}\}$$

$D=\{NDVI \text{ Aprox}\}$

$Q=\{\text{Nivel de Radianza Banda Espectral 5, Nivel de Radianza Banda Espectral 4, NDVI Aprox}\}$

$V = [-1.0, -0.9, -0.8, -0.7, \dots, 0.7, 0.8, 0.9, 1.0]$ que corresponde a cada una de las clases $[21, 19, 18, 17, \dots, 12, 11, 1, 2, 3, 4, \dots, 9, 10]$ según los NDVI Aprox $f: U \times Q \rightarrow V$ función tal que $f(x, q) \in V$, para $x \in U$, llamada función de información.

$f(\text{Objeto}, \text{atributo})$

$f(D1, \text{Banda Espectral 5})=4$, $f(D1, \text{Banda espectral 4})=75$, $f(D1, NDVI Aprox)=-0.9, \dots$

3.2.1 Discernibilidad. Se dice que x está relacionado con y , se escribe: $xI_P y$, si para todo atributo $q \in P$, donde P es un subconjunto de atributos de V , se cumple:

$f(x1, q) = f(x2, q)$

Entonces para un conjunto P de atributos, $P=\{\text{Banda espectral 5, Banda espectral 4}\}$ se tienen los elementos discernibles e indiscernibles:

$I_p(D1)=\{D1\}$	$I_p(D14)=\{D14\}$
$I_p(D2)=\{D2\}$	$I_p(D15)=\{D15\}$
$I_p(D3)=\{D3\}$	$I_p(D16)=\{D16\}$
$I_p(D4)=\{D4\}$	$I_p(D17)=\{D17\}$
$I_p(D5)=\{D5\}$	$I_p(D18)=\{D18\}$
$I_p(D6)=\{D6\}$	$I_p(D19)=\{D19\}$
$I_p(D7)=\{D7\}$	$I_p(D20)=\{D20\}$
$I_p(D8)=\{D8\}$	$I_p(D21)=\{D21\}$
$I_p(D9)=\{D9\}$	$I_p(D22)=\{D22\}$
$I_p(D10)=\{D10\}$	$I_p(D23)=\{D23\}$
$I_p(D11)=\{D11\}$	$I_p(D24)=\{D24\}$
$I_p(D12)=\{D12\}$	$I_p(D25)=\{D25\}$
$I_p(D13)=\{D13\}$	

En el caso anterior no hay elementos indiscernibles, esto por las características del cálculo del NDVI, es decir que los dos atributos son necesarios para poder definir la clase a la que pertenece, de esta forma no hay aproximaciones ni es necesario calcular las dependencias y reducciones, para el conjunto de atributos

1. $P1=\{\text{Banda Espectral 5}\}$
2. $P2=\{\text{Banda Espectral 4}\}$
3. $P=\{\text{Banda Espectral 5, Banda Espectral 4}\}$

Se observa que el *NDVI Aprox*, requiere de los dos atributos $P1$ y $P2$

3.3 Obtener las reglas de decisión

Las reglas de decisión como se mencionó anteriormente son el tipo Si.. ENTONCES, las cuales se construyen a partir de los atributos P1 y P2 es decir P, así, para $P = \{\text{Banda Espectral 5, Banda Espectral 4}\}$, que es el conjunto formado por los dos atributos, se tienen las reglas de decisión así:

- SI $f(x, P1)=4$ y $f(x, P2)=7$ ENTONCES
 $f(x, \text{Estado}) = -0.898734177 = \text{clase 19}$
- SI $f(x, P1)=21$ y $f(x, P2)=71$ ENTONCES
 $f(x, \text{Estado}) = -0.543478261 = \text{clase 16}$
- SI $f(x, P1)=39$ y $f(x, P2)=0$ ENTONCES
 $f(x, \text{Estado}) = 1 = \text{clase 1}$
- SI $f(x, P1)=43$ y $f(x, P2)=103$ ENTONCES
 $f(x, \text{Estado}) = -0.410958904 = \text{clase 15}$
- SI $f(x, P1)=50$ y $f(x, P2)=99$ ENTONCES
 $f(x, \text{Estado}) = -0.32885906 = \text{clase 14}$
- SI $f(x, P1)=54$ y $f(x, P2)=83$ ENTONCES
 $f(x, \text{Estado}) = -0.211678832 = \text{clase 13}$
- SI $f(x, P1)=62$ y $f(x, P2)=79$ ENTONCES
 $f(x, \text{Estado}) = -0.120567376 = \text{clase 12}$
- SI $f(x, P1)=75$ y $f(x, P2)=99$ ENTONCES
 $f(x, \text{Estado}) = -0.137931034 = \text{clase 12}$
- SI $f(x, P1)=83$ y $f(x, P2)=103$ ENTONCES
 $f(x, \text{Estado}) = -0.107526882 = \text{clase 12}$
- SI $f(x, P1)=96$ y $f(x, P2)=63$ ENTONCES
 $f(x, \text{Estado}) = 0.20754717 = \text{clase 13}$
- SI $f(x, P1)=104$ y $f(x, P2)=107$ ENTONCES
 $f(x, \text{Estado}) = -0.014218009 = \text{clase 11}$
- SI $f(x, P1)=118$ y $f(x, P2)=59$ ENTONCES
 $f(x, \text{Estado}) = 0.333333333 = \text{clase 3}$
- SI $f(x, P1)=121$ y $f(x, P2)=136$ ENTONCES
 $f(x, \text{Estado}) = -0.058365759 = \text{clase 11}$
- SI $f(x, P1)=135$ y $f(x, P2)=67$ ENTONCES

- $f(x, Estado) = 0.336633663 = \text{clase 3}$
- SI $f(x, P1)=147$ y $f(x, P2)=47$ ENTONCES
 $f(x, Estado) = 0.515463918 = \text{clase 5}$
- SI $f(x, P1)=158$ y $f(x, P2)=144$ ENTONCES
 $f(x, Estado) = 0.046357616 = \text{clase 0}$
- SI $f(x, P1)=164$ y $f(x, P2)=148$ ENTONCES
 $f(x, Estado) = 0.051282051 = \text{clase 0}$
- SI $f(x, P1)=172$ y $f(x, P2)=148$ ENTONCES
 $f(x, Estado) = 0.075 = \text{clase 0}$
- SI $f(x, P1)=185$ y $f(x, P2)=196$ ENTONCES
 $f(x, Estado) = -0.028871391 = \text{clase 11}$
- SI $f(x, P1)=198$ y $f(x, P2)=192$ ENTONCES
 $f(x, Estado) = 0.015384615 = \text{clase 11}$
- SI $f(x, P1)=206$ y $f(x, P2)=176$ ENTONCES
 $f(x, Estado) = 0.078534031 = \text{clase 11}$
- SI $f(x, P1)=214$ y $f(x, P2)=188$ ENTONCES
 $f(x, Estado) = 0.064676617 = \text{clase 11}$
- SI $f(x, P1)=220$ y $f(x, P2)=204$ ENTONCES
 $f(x, Estado) = 0.037735849 = \text{clase 11}$
- SI $f(x, P1)=233$ y $f(x, P2)=204$ ENTONCES
 $f(x, Estado) = 0.066361556 = \text{clase 11}$
- SI $f(x, P1)=249$ y $f(x, P2)=204$ ENTONCES
 $f(x, Estado) = 0.099337748 = \text{clase 11}$

Las anteriores reglas son determinísticas, con base en ella se diseña el algoritmo de clasificación, el cual de forma experimental se implementó en dos etapas, en la primera se establece la forma como se recupera la información de las imágenes satelitales, desde el traslado de las imágenes a clasificar hasta la entrega de un vector con los niveles de radianza de cada una de las bandas espectrales por cada píxel que compone la imagen y una segunda parte donde se asigna un falso color que permita dar una apariencia similar a la vegetación dependiendo de la radianza encontrada.

4. Conclusiones

- La técnica *Rough Set* es aplicable en conjuntos de datos donde sean poco diferenciables sus elementos en función de sus atributos o características, es decir, que no sea posible clasificar o agrupar en subconjuntos; es amplia y diversa la aplicación que ha tenido esta técnica, lo que demuestra su potencialidad y factibilidad para obtener buenos resultados al clasificar imágenes satelitales.
- El proceso de clasificación de imágenes satelitales se constituye en un campo en el cual se puede aprovechar las ventajas de la técnica *Rough Set*, ya que los niveles digitales de una imagen satelital para un espectro electromagnético determinado, presentan vaguedad e incertidumbre al ubicar los objetos dentro de un conjunto específico, dada las características de sus atributos.
- Desde el punto de vista de los recursos necesarios para la aplicación de *Rough Set* en la clasificación de imágenes satelitales, ésta tiene un alto costo computacional dado las características propias del algoritmo y del tamaño en bytes de las imágenes satelitales, pero igual es abordable dados las grandes capacidades de cómputo y de avances tecnológicos a nivel de hardware, software y procesamiento en la *grid* disponibles hoy en día.
- Por lo anteriormente expuesto, se encuentra un potencial elevado al buscar el aprovechamiento de esta técnica al clasificar elementos dentro de una superficie terrestre captada como niveles de *radianza* en una imagen satelital, ya se apoyándose en una supervisión en el proceso o no.

Bibliografía

- ADRIAANS, Pieter (2002). Production control. In: KLOSGEN, Willi and ZYTKOW, Jan M. (Eds.), Handbook of data mining and knowledge discovery. New York (USA): Oxford University Press. 1064 p. ISBN: 978-0195118315
- ADRIAANS, Pieter and ZANTINGE, Dolf (1996). Data Mining. New York (USA): Addison Wesley. 176 p. ISBN: 978-0201403800
- ALDRIDGE, Colin Hamilton (1998) A Theory Of Empirical Spatial Knowledge Supporting Rough Set Based Knowledge Discovery in Geographic Databases. PhD Thesis, Dunedin (New Zealand): University of Otago.
- ARCO, Leticia and BELLO, Rafael (2006). Rough Text: a Rough Approach in Text Processing. In: ISFUROS 2006 (5-8/12/2006), Santa Clara (Cuba): Department of Computer Science, Central University of Las Villas. Proceeding of the International Symposium on Fuzzy and Rough Sets, ISFUROS 2006.
- BONALDI, E.L.; BORGES DA SILVA, E.; LAMBERT TORRES, G.; OLIVEIRA, L. and ASSUNCAO, F.O. (2002). Using Rough Set Techniques as Fault Diagnosis Classifier for Induction Motors [on line]. In: 28th Annual Conference of the Industrial Electronics Society, IECON 02 (5-8/11/2002). Seville (Spain): IEEE. Proceedings of the 2002 28th Annual Conference of the Industrial Electronics Society, IECON 2002, Vol. 4, p. 3383 - 3388 vol. 4 ISBN: 0-7803-7474-6 <http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1182941> [consult: 18/07/2011]
- BOSQUE SENDRA, Joaquín (1997). Sistemas de información geográfica. Madrid (España): Rialp, 451 p. ISBN: 84-321-3154-7
- BOSQUE SENDRA, Joaquín y GARCÍA, Rosa C. (2000). El uso de los sistemas de información geográfica en al planificación territorial. En: Anales de Geografía, No. 20. Madrid (España): Universidad Complutense. p. 49-67. ISSN: 0211-9803
- BUSINESS IMAGE GROUP Y SPOT IMAGE (1999). Imágenes de satélite: Una Guía Objetiva. París (Francia): Centre National d'Etudes Spatiales (CNES).
- CABALLERO MOTA, Yailé; BELLO PÉREZ, Rafael; GARCÍA LORENZO, María Matilde; SUÁREZ, Yadiika; Rivero, Inclán; RODRÍGUEZ VALLEJO, Lester; MADERA QUINTANA, Julio César (2007a). Uso de los conjuntos aproximados para el tratamiento de los datos. En: X Congreso de la Sociedad Cubana de Matemática y Computación, Compumat 2007 (nov.), Holguín (Cuba): Sociedad Cubana de Matemáticas y Computación. Boletín de la Sociedad Cubana de Matemáticas y Computación, Vol. 4, No. 2, ISSN: 1728-6042.
- CABALLERO, Yailé; ÁLVAREZ Delia, BALTÁ, Analay; BELLO, Rafael y GARCÍA, María. (2007b). Un nuevo algoritmo de selección de rasgos basado en la teoría de los conjuntos aproximados [en línea]. En: Revista Facultad de Ingeniería, No. 41 (sep.). Medellín (Colombia): Universidad de Antioquia, p. 132-144. ISSN: 0120-6230. <<http://redalyc.uaemex.mx/pdf/430/43004111.pdf>> consulta: 23/08/2011]
- CAICEDO CAICEDO, Julio César y PÉREZ CASTILLO, José Nelson (2010). Servicio web inteligente para la clasificación de imágenes digitales utilizando conjuntos aproximados [en línea]. En: Ingeniería e Investigación. Vol. 30, No. 1 (abr.) Bogotá (Colombia): Universidad nacional de Colombia. p. 45-51. ISSN: 0120-5609. <<http://redalyc.uaemex.mx/src/inicio/ArtPdfRed.jsp?iCve=64312498009>> [consulta: 14/08/2011]
- CÁRDENAS, Beitmantt; CABALLERO, Yailé y BELLO, Rafael (2007). La Teoría de los Conjuntos Aproximados y las Técnicas de Bootstrap para la edición de conjuntos de entrenamiento. Su aplicación en el pronóstico meteorológico [en línea]. En: Avances en Sistemas e Informática, Vol. 4, No. 3 (dic.). Medellín (Colombia): Universidad Nacional de Colombia, p. 165-170. ISSN 16577663 <<http://pisis.unalmed.edu.co/avances/archivos/ediciones/Edicion%20Avances%202007%203/18.pdf>> [consulta: 14/07/2011]
- CASI, CANADIAN AERONAUTICS AND SPACE INSTITUTE (2010). 30th. Canadian Symposium on Remote Sensing: Bridging excellence (on line). Canadian Journal of Remote Sensing, Vol. 36, Suppl. 2. Kanata (Ontario, Canada): Canadian Aeronautics and Space Institute. ISSN: 1712-7971.

- CHOUDHARIL, Ajay and NANDI, G.C. (2005). Indiscernibility criterion based on Rough sets in Feature selection and Detection of Landmines [on line]. In: IEEE International Conference on Granular Computing 2005, (27/07/2005), Beijing (China): IEEE. Proceedings IEEE International Conference on Granular Computing 2005, Vol. 2, p. 413-416. ISBN: 0-7803-9017-2. <<http://ieeexplore.ieee.org/Xplore/login.jsp?url=http%3A%2F%2Fieeexplore.ieee.org%2Fiel5%2F10381%2F33011%2F01547324.pdf%3Farnumber%3D1547324&authDecision=-203>> [consult: 23/07/2011]
- CHUVIECO SALINERO, Emilio (2002). Teledetección ambiental: La observación de la tierra desde el espacio. Barcelona (España): Ariel Ciencia. 586 p. ISBN: 978-8434434981.
- CONAF, UNIVERSIDAD DE CHILE (1983). Proyecto Teledetección para manejo forestal dinámico Temford. Informe final. CONAF. Santiago de Chile (Chile): CONAF. 76 p.
- DELGADO INGA, Omar (2009). Nuevos sensores de teledetección [en línea]. En: Geomática, Revista de la Universidad del Azuay No. 49 (ago.). Cuenca (Ecuador): Universidad de Azuay. p. 53-80. ISSN: 13902849 <http://www.uazuay.edu.ec/bibliotecas/publicaciones/UV-49_Geomática.pdf> [consulta: 23/07/2011]
- DÍAZ MARTÍNEZ, Zuleyka; FERNÁNDEZ MENÉNDEZ, José y SEGOVIA VARGAS, María Jesús (2004). Sistemas de Inducción de Reglas y Árboles de Decisión Aplicados a la Predicción de Insolvencias en Empresas Aseguradoras [en línea]. Madrid (España): Universidad Complutense de Madrid. 13 p. <<http://eprints.ucm.es/6833/1/0409.pdf>> [consulta: 25/07/2011]
- DÜNTSCH, Ivo and GEDIGA, Günther (2000). Rough set data analysis: A road to non-invasive knowledge discovery. Bangor (UK): Methods Publishers. ISBN: 190328001X
- ESRI CHILE (s.f.). Landsat 7 ETM+ [en línea]. Santiago (Chile): ESRI Chile S.A. <<http://www.esri-chile.com/biblioteca/landsat7etm.pdf>> [consulta: 23/07/2011]
- FAO (1980). Metodología provisional para la evaluación de la degradación de los suelos. Roma (Italia): FAO, 86 p.
- FELICÍSIMO, Ángel M. (2001). Glosario de términos usados en el trabajo con sistemas de información geográfica [en línea]. Oviedo (España): Escuela Técnica Superior de Ingenieros de Minas, Universidad de Oviedo. <<http://www.etsimo.uniovi.es/~feli/index2.html>> [consulta: 26/07/2011].
- FERNÁNDEZ COPPEL, Ignacio Alonso y HERRERO LLORENTE, Eliécer (2001). El satélite Landsat: Análisis visual de imágenes obtenidas del sensor ETM + satélite Landsat [en línea]. Valladolid (España): Universidad de Valladolid. 37 p. <<http://www.caece.edu.ar/tea/Apuntes/landsat-analisis-visual.pdf>> [consulta: 22/07/2011]
- GATES, David M. (2003). Biophysical Ecology, New York (NY, USA): Dover Publications, 635 p. ISBN: 978-0486428840.
- GEOSERVICE PERÚ (s.f.). Información Técnica Resumida – Quickbird II [en línea]. Lima (Perú): GeoService Perú. 3 p. <http://www.gsp Peru.com/pdf/res_quickbird.pdf> [consulta: 23/07/2011]
- GLCF (2011). Global Land Cover Facility [on line]. College Park (Maryland, USA): University of Maryland - NASA. <<http://www.landcover.org/>> [consult: 10/11/2011]
- GÓMEZ A, Héctor F. (2006) Clasificación de imágenes multispectrales Landsat TM por medio de redes neuronales no supervisadas [en línea]. Tesis doctoral (Doctorado en inteligencia artificial: Redes Neuronales). Loja (Ecuador): Universidad Nacional de Estudios a Distancia, UNED - Universidad Técnica Particular de Loja. 34 p. <http://sig.utpl.edu.ec/sigutpl/Staftpro/realidad/Clasificacin_imgenes_multispectrales.pdf> [consulta: 23/07/2011]
- GONZÁLEZ ALONSO, F. y CUEVAS GOZALO, J.M. (1982). Los satélites de recursos naturales y sus aplicaciones en el campo forestal. En: Monografías INIA, No. 34. Madrid (España): Instituto Nacional de Investigaciones Agrarias, INIA. 46 p. ISSN: 0210-3354
- GROSSMAN, Robert; KASIF, Simon; MOORE, Reagan; ROCKE, David and ULLMAN, Jeff (1998). Data mining research: opportunities and challenges [on line]. In: Three NSF workshops on mining large, massive, and distributed data (01/1998). Proceedings of 3 NSF. New York (USA): AAAI Press. 11 p. <<http://pubs.rgrossman.com/dl/misc-001.pdf>> [consult: 22/07/2011]
- HOR, Ching-Lai and CROSSLEY, Peter A. (2005) Knowledge Extraction from Intelligent Electronic Devices. In: PETERS, James F. and SKOWRON, Andrzej (Ed.). Transactions on Rough Sets III. Berlin Heidelberg (Germany): Springer-Verlag, p. 82-111. ISBN 978-3-642-18301-0.

- JENSEN, John R. (1995). *Introductory digital image processing: a remote sensing perspective*. 2 ed. Upper Saddle River (NJ, USA): Prentice Hall Series. 316 p. ISBN: 01232058405.
- KOHAVI, R. and FRASCA, B. (1994). Useful feature subsets and Rough set Reducts. In: *Third International Workshop on Rough Sets and Soft Computing (RSSC'94)*. (10-12/11/1994). San José (CA, USA). *Proceedings of RSSC'94*. p. 320-323.
- KOMOROWSKI, J.; PAWLAK, Z.; POLKOWSKI, L. and SKOWRON, A. (1999). A Rough sets: a tutorial. In: PAL, S.; SKOWRON, A. *Rough Fuzzy Hybridization: A new trend in Decision Making*. Singapore: Springer-Verlag. p. 3-98
- LI, Maozhen; YU, Bin; HUANG, Chang; SONG, Yong-Hua and RANA, O.F. (2006). Service Match-making with Rough Sets. In: *Sixth IEEE International Symposium on Cluster Computing and the Grid (CCGRID'06)*. (16-19/05/2006). Singapore: IEEE Computer Society. *Proceedings of CCGRID'06*. 8 p. ISBN: 0-7695-2585-7.
- LILLESAND, Thomas and M. KIEFER, Ralph W. (1999). *Remote Sensing and Image Interpretation*. 4 ed. Hoboken (NJ, USA): John Wiley & Sons Inc. 736 p. ISBN: 0471255157
- NASA (s.f.). *Landsat 7 Science Data Users Handbook* [on line]. Washington (USA): National Aeronautics and Space Administration, NASA. 186 p. <http://landsathandbook.gsfc.nasa.gov/pdfs/Landsat7_Handbook.pdf> [consult: 23/07/2011]
- NAVARRO BONILLA, Diego y ESTEBAN NAVARRO, Miguel Ángel (2003) (eds.). *I Seminario sobre Gestión del Conocimiento y Servicios de Inteligencia en el siglo XXI*. Madrid (España): Universidad Carlos III de Madrid. 231 p.
- NGO, Chi Lang and NGUYEN, Hung Son (2005). A method of web search result clustering based on rough sets. [on line]. In: *International Conference on Web Intelligence, WI 2005*. (19-22/09/2005), Compiegne (France): IEEE, WIC, ACM. *Proceedings of Web Intelligence*. IEEE Computer Society, p. 673-679. ISBN 0-7695-2415-X <http://logic.mimuw.edu.pl/publikacje/WI_Nguyen_Son.pdf> [consult: 22/07/2011]
- ØHRN, Aleksander (1999). *Discernibility and Rough Sets in Medicine: Tools and Applications* [on line]. PhD thesis (doktor ingeniør). Trondheim (Norway): Department of Computer and Information Science, Norwegian University of Science and Technology. 239 p. ISBN: 82-7984-014-1. <ntnu.diva-portal.org/smash/get/diva2:125243/FULLTEXT01> [consult: 11/08/2011]
- PAWLAK, Zdzislaw (1982) *Rough Sets*. In: *International Journal of Computer and Information Sciences*, Vol 11, No. 5, (Oct.). ISSN: 0091-7036. Warsaw (Poland): Inst. of Computer Sci., Polish Acad. of Sci. p. 341-356.
- PAWLAK, Zdzislaw (1991). *Rough Sets: Theoretical Aspects of Reasoning about Data*. Boston (USA): Kluwer Academic Publishers. 252 p. ISBN: 978-0792314721.
- PAWLAK, Zdzislaw (2002). *Rough sets and their applications* [on line]. In: *Journal of Telecommunications and Information Technology*, No. 3. Miedzeszyn (Poland): National Institute of Telecommunications. p. 7-10. <<http://www.nit.eu/czasopisma/JTIT/2002/3/7.pdf>> [consult: 11/11/2011].
- PETERS, J.F.; SURAJ, Z.; SHAN, S.; RAMANNA, S.; PEDRYCZ, W. and PIZZI, N. (2003). Classification of meteorological volumetric radar data using rough set methods. *Pattern Recognition Letters* Vol. 24, No. 6 (mar.). New York (NY, USA): Elsevier Science Inc. p. 911-920. ISSN: 0167-8655
- PREDKI, B.; WILK, Sz (1999). *Rough Set Based Data Exploration Using ROSE System*. In: Z.W.Ras, A.Skowron, eds. *Foundations of Intelligent Systems, Lecture Notes in Artificial Intelligence*, vol. 1609. Springer-Verlag, Berlin.
- PREDKI, B; SLOWINSKI, R.; STEFANOWSKI, J.; SUSMAGA, R. and ROSE (1998). *Software Implementation of the Rough Set Theory*. In: POLKOWSKI, L. and SKOWRON, A (eds) (1998). *Rough Sets and Current Trends in Computing, Lecture Notes in Artificial Intelligence*, vol. 1424. Berlin (Germany): Springer-Verlag.
- RICHARDS, John A. and JIA, Xiuping (2005). *Remote Sensing Digital Image Analysis. An Introduction*. 4 ed. Berlin (Germany): Springer-Verlag. 464 p. ISBN: 978-3540251286.
- SEGOVIA VARGAS, María Jesús. (2005) *Análisis De La Solvencia En Entidades Aseguradoras Mediante La Metodología Rough Set*. Ed. Civitas. ISBN: 978-84-470-2316-5.
- SEGOVIA VARGAS, María Jesús; GIL FANA, J.A.; HERAS MARTÍNEZ, A.; VILAR ZANÓN, J.L. (s.f. b). *Predicción de insolvencias con el método Rough Set*. Dpto. de Economía Financiera

- y Contabilidad I, Facultad de C. Económicas y Empresariales. Madrid (España): Universidad Complutense de Madrid.
- SEGOVIA VARGAS, María Jesús; GIL FANA, J.A; HERAS MARTÍNEZ, A.; VILAR ZANÓN, J.L. (s.f. a). La metodología Rough Set frente al Análisis Discriminante en los problemas de clasificación multiaatributo. Departamento de Economía Financiera y Contabilidad I, Facultad de C. Económicas y Empresariales. Madrid (España): Universidad Complutense de Madrid.
- SPERANZA, F. y ZERDA, H. (2005). Potencialidad de los índices de vegetación para la discriminación de coberturas forestales. Proyecto Picto 12931, INTA Manfredi. Santiago del Estero (Córdoba, Argentina): Facultad de Ciencias Forestales, UNSE.
- WAKAKI, Toshiko; ITAKURA, Hiroyuki and TAMURA, Masaki (2004). Rough set-Aided Feature Selection for Automatic Web-Page Classification. In: Proceeding of the International Conference on Intelligence IEEE/WIC/ACM (WI'04). Washington (USA): IEEE Computer Society. ISBN: 0-7695-2100-2
- XIAN MING, Huang; JI-KAI, Yi and YAN-HONG, Zhang (2003). A method of constructing fuzzy neural network based on rough set theory. In: International Conference on Machine Learning and Cybernetics. (02-05/11/2003). Xi'an (China): IEEE. Proceedings Vol. 3, p. 1723-1728. ISBN: 0-7803-7865-2.
- YAO, Y.Y.; WONG, Y.S. and BUTZ, C.J. (1999). On Information-Theoretic Measures of attribute important. In: ZHONG, Ning and ZHOU, Lizhu (Eds.) (1999). Methodologies for Knowledge Discovery and Data Mining: Third Pacific-Asia Conference, PAKDD'99, (26-28/04/1999). Beijing (China): Springer. ISBN: 978-3540658665
- YING, Sai; PEIYAO, Nie and DONGSHENG, Chu (2005). A Model of Granular Computing Based on Rough Set Theory. In: HU, Xiaohua; LIU, Qing; SKOWRON, Andrzej; YOUNG LIN, Tsau; YAGER, Ronald R. and ZHANG, Bo (2005). IEEE International Conference on Granular Computing. (25-27/07/2005). Beijing (China): IEEE Computational Intelligence Society. Vol. 1, p. 233-236. ISBN: 0-7803-9017-2.
- YONG Xu; JING-YU Yang; ZHONG, Jin and YU-JIE, Zheng (2006). Local Correlation Classification and Its Application to Face Recognition Across Illumination. In: Fifth International Conference on Machine Learning and Cybernetics (13-16/08/2006). Dalian (Liaoning, China): IEEE. Proceedings p. 3277-3281. ISBN: 1-4244-0061-9.
- YUN, Gao and YUANBIN, Hou (2004). Application of Rough Set Theory on System Modeling. In: Fifth World Congress on Intelligent Control and Automation WCICA 2004. (15-19/06/2004). Hangzhou (China): IEEE. Vol. 3, p. 2352-2354. ISBN: 0-7803-8273-0.
- ZHONG, Ning; DONG, Juzhen and OHSUGA, Setsuo (2001) Using Rough sets with heuristics for feature selection [on line]. In: Journal of Intelligent Information Systems. Vol. 16, No. 3, p. 199-214. ISSN: 1573-7675 <<http://www.springerlink.com/content/q096486411525j63/>> [consult: 12/08/2011]
- ZOPOUNIDIS, Constantin and DIMITRAS, Agustinos I. (1998). Multicriteria decision methods aid for the prediction of business failure. Berlin (Germany): Springer. 171 p. ISBN: 978-0792349006.