Introducción a la Teoría de los Conjuntos aproximados: conceptos y aplicaciones.
(Rough Sets Theory, RST)

Dr. Rafael Bello Pérez
Departamento de Ciencia de la Computación
Universidad Central de Las Villas
Cuba
Email: rbellop@uclv.edu.cu

MOTIVACION

Análisis de datos

Incertidumbre

- Incertidumbre
- Imprecisión
- Inconsistencia
- Vaguedad
- Incompletitud

Aprendizaje

- Evaluación de CE
- Selección de rasgos
- Relación de atributos
- Aprendizaje inductivo

Teoría de los conjuntos aproximados

Introducción

- **Rough set theory** fue desarrollada por Zdzislaw Pawlak in the early 1980's.
- Publicaciones Representativas:
  - Z. Pawlak, "Rough Sets", *International Journal of Computer and Information Sciences*, Vol.11, 341-356 (1982).
  - Z. Pawlak, *Rough Sets - Theoretical Aspect of Reasoning about Data*, Kluwer Academic Pubilishers (1991).

Valoraciones sobre la teoría:

Esta teoría ha mostrado ser una excelente herramienta matemática para el análisis de datos, Salvatore Greco, University of Catania, Italia, 2001.

La teoría de los conjuntos aproximados ha mostrado ofrecer una exitosa base teórica para la solución de muchos problemas dentro del descubrimiento de conocimiento, Torulf Mollestad y Jan Komorowski, Norwegian University of Science and Technology, Norway, 1999

Valoraciones sobre la teoría:

La teoría de los conjuntos aproximados es, potencialmente, una importante herramienta en el análisis de datos con importantes aplicaciones en la minería de datos y el descubrimiento de conocimiento, W. Koczkodaj, Laurentian University, Canada, M. Orlowski, University of Technology, Australia, y V. Marek, University of Kentucky, USA, 1998.

El uso de la teoría de los conjuntos aproximados para el descubrimiento de conocimiento y minar reglas es ampliamente reconocido, Sankar K. Pal

Representación de la RST:

RST = Sistema de información + Relación de separabilidad.

Relación de equivalencia ⇒ RST clásica.

Relación de tolerancia o similaridad ⇒ RST extendida

La relación de separabilidad genera una granulación del

Sistema de Información

Definición: (Sistema de Información)

Sea un conjunto de atributos A=$\{$A1, A2,...,An$\}$ y un conjunto U no vacío llamado universo de ejemplos (objetos, entidades, situaciones o estados, etc.) descritos usando los atributos Ai. Al par (U,A) se le denomina Sistema de información.

Definición: (Sistema de decisión)

Si a cada elemento de U se le agrega un nuevo atributo d llamado decisión indicando la decisión tomada en ese estado o situación entonces se obtiene un Sistema de decisión (U, A$\cup\{$d$\}$, donde d$\notin$A).

EJEMPLO DE SISTEMA DE DECISION

Paciente	Dolor de cabeza	Dolor muscular	Temperatura	Gripe
P1	no	si	alta	Si
P2	si	no	alta	Si
P3	si	si	muy alta	Si
P4	no	si	normal	No
P5	si	no	alta	No
P6	no	si	muy alta	Si

Sistemas de Información /Tablas

	Age	LEMS
x 1	16-30	50
x2	16-30	0
x3	31-45	1-25
x4	31-45	1-25
x5	46-60	26-49
x6	16-30	26-49
x7	46-60	26-49

- IS is a pair (*U*, *A*)
- *U* is a non-empty finite set of objects.
- *A* is a non-empty finite set of attributes such that $a: U \to V_a$ for every $a \in A$.
- V_a is called the value set of *a*.

Sistemas de Decisión/Tablas

	Age	LEMS	Walk
x 1 yes	16-30	50	
x2	16-30	0	no
x3	31-45	1-25	no
x4	31-45	1-25	yes
x5	46-60	26-49	no
x6	16-30	26-49	yes
x7	46-60	26-49	no

- DS: $T = (U, A \cup \{d\})$
- $d \notin A$ is the *decision* attribute (instead of one we can consider more decision attributes).
- The elements of *A* are called the *condition* attributes.

RELACION DE SEPARABILIDAD

Se dice que un atributo $Ai \in A$ separa o distingue un objeto x de otro y si y solo si:

(R1) $Separa(Ai,x,y) \Leftrightarrow f(x,Ai) \neq f(y,Ai)$

Matriz de separación MA, la cual es una matriz $|U|x|U|$, cada entrada $MA(x,y) \subseteq A$ contiene el conjunto de atributos que distinguen los elementos x e y de U:

Ejemplo de Matriz de separación:

	P1	P2	P3	P4	P5	P6
P1		{1, 2}	{1, 3 }	{3, 4}	{1, 2, 4}	{3}
P2			{2, 3}	{1, 2, 3, 4}	{4}	{1, 2, 3}
P3				{1, 3, 4}	{2, 3, 4}	{1}
P4					{1, 2, 3}	{3, 4}
P4						{1, 2, 3, 4}
P6						

Ejemplo de Matriz de Separación

No	a	b	c	d
u1	*a0*	*b1*	*c1*	*y*
u2	*a1*	*b1*	*c0*	*n*
u3	*a0*	*b2*	*c1*	*n*
u4	*a1*	*b1*	*c1*	*y*

In order to discern equivalence classes of the decision attribute *d,* to preserve conditions described by the discernibility matrix for this table

	u1	u2	u3
u2	a,c		
u3	b	λ	
u4	λ	c	a,b

$C = \{a, b, c\}$

$D = \{d\}$

$(a \vee c) \wedge b \wedge c \wedge (a \vee b)$

$= b \wedge c$

Reduct = $\{b, c\}$

OTRAS RELACIONES DE SEPARABILIDAD

Relación de separabilidad para sistemas de información incompletos:

Valores omitidos o desconocidos, denotados por asterisco (*); es decir, $V_i = V_i \cup \{*\}$.

(R2) $\text{Separa}(A_i,x,y) \Leftrightarrow f(x,A_i) \neq f(y,A_i)$ and $f(x,A_i) \neq *$ and $f(y,A_i) \neq *$

Relación de separabilidad entre objetos basada en similaridad:

(R3) $\text{Separa}(A_i,x,y) \Leftrightarrow | f(x,A_i) - f(y,A_i) | > \varepsilon$

Consistencia

Sea la función δB:

$\delta\, B(x)=\{\ v \in Vd : \exists y \in IB(x)$ tales que

$d(y)=v\ \}$

Un sistema de decisión es consistente, si y solo si, el conjunto $\delta B(x)$ es un conjunto unitario para todo $x \in U$.

Definición: (Relación de inseparabilidad):

- A cada subconjunto de atributos B de A ($B \subseteq A$) esta asociada una relación binaria de inseparabilidad denotada por IB, la cual es el conjunto de pares de objetos que son inseparables uno de otros por esa relación.
- $IB=\{ (x,y) \in UxU \ f(x,Ai)=f(y,Ai)$ para todo $Ai \in B\}$
- Una relación que es simétrica, reflexiva y transitiva es una *relación de equivalencia.*

Relación de equivalencia

- The equivalence relation

  A binary relation $R \subseteq X \times X$ which is reflexive (*xRx* for any object *x*) ,

  symmetric (if *xRy* then *yRx*), and

  transitive (if *xRy* and *yRz* then *xRz*).

- The equivalence class $[x]_R$ of an element $x \in X$

  consists of all objects $y \in X$ such that *xRy*.

Indiscernibilidad

- Let *IS = (U, A)* be an information system, then with any $B \subseteq A$ there is an associated equivalence relation:

$$IND_{IS}(B) = \{(x, x') \in U^2 \mid \forall a \in B, a(x) = a(x')\}$$

  where $IND_{IS}(B)$ is called the *B-indiscernibility relation.*
- If $(x, x') \in IND_{IS}(B),$ then objects *x* and *x'* are indiscernible from each other by attributes from *B.*
- The equivalence classes of the *B-indiscernibility relation* are denoted by $[x]_B$.

Un Ejemplo de Indiscernibilidad

	Age	LEMS	Walk
x 1 yes	16-30	50	
x2	16-30	0	no
x3	31-45	1-25	no
x4	31-45	1-25	yes
x5	46-60	26-49	no
x6	16-30	26-49	yes
x7	46-60	26-49	no

- The non-empty subsets of the condition attributes are *{Age}*, *{LEMS}*, and *{Age, LEMS}*.
- *IND({Age}) = {{x1,x2,x6}, {x3,x4}, {x5,x7}}*
- *IND({LEMS}) = {{x1}, {x2}, {x3,x4}, {x5,x6,x7}}*
- *IND({Age,LEMS}) = {{x1}, {x2}, {x3,x4}, {x5,x7}, {x6}}*

Observaciones

- Una relación de equivalencia induce una **granulación** del universo que es una **partición**.
- Las particiones se pueden usar para construir nuevos subconjuntos de el universo.
- Los Subconjuntos de más interés son aquellos que tienen el mismo valor del atributo decisión.

Aproximaciones de un conjunto

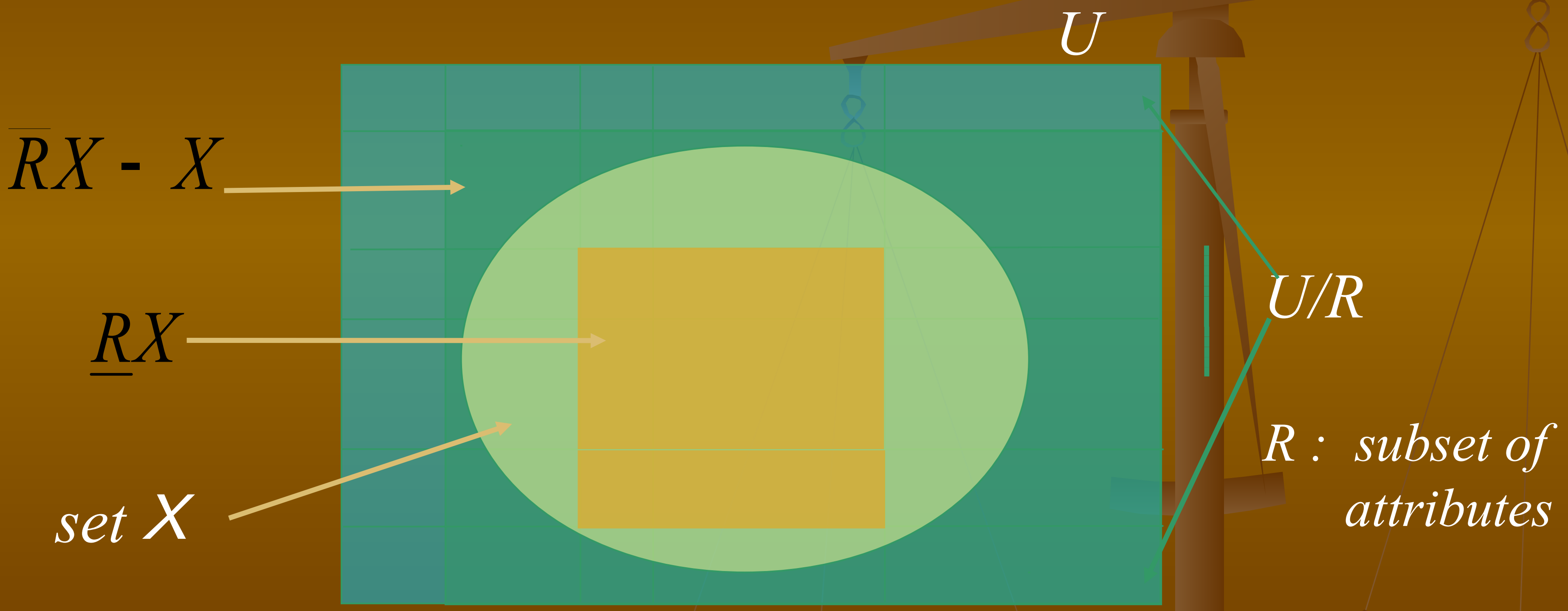
Sea $B \subseteq A$ una relación de equivalencia:

$Bi(X) = \{x \in U \mid B(x) \subseteq X\}$ □ aproximación inferior de X

$Bs(X) = \{x \in U \mid B(x) \cap X \neq \phi\}$ □ aproximación superior de X

$BNB(X) = Bs(X) - Bi(X)$ □ frontera de X

Aproximaciones Inferior & Superior



Propiedades de los conjuntos aproximados son:

a) $Bi(X) \subseteq X \subseteq Bs(X)$

b) $Bi(\phi) = Bs(\phi) = \phi$

c) $Bi(U) = Bs(U) = U$

d) $Bs(X \cup Y) = Bs(X) \cup Bs(Y)$

e) $Bi(X \cap Y) = Bi(X) \cap Bi(Y)$

f) $X \subseteq Y \Rightarrow Bi(X) \subseteq Bi(Y)$ y $Bs(X) \subseteq Bs(Y)$

g) $Bi(X) = U\text{-}Bs(U\text{-}X)$, propiedad de complementariedad.

Regiones

- la región positiva: POS(X)= Bi(X).
- la región limite: BND(X)= BNB(X).
- la región negativa: NEG(X)= U-Bs(X).
- Los conjuntos Bi(X), Bs(X), POS(X), BND(X) y NEG(X) son las nociones más importantes de la teoría de conjuntos aproximados.

Un Ejemplo de Aproximación

	Age	LEMS	Walk
x1	16-30	50	yes
x2	16-30	0	no
x3	31-45	1-25	no
x4	31-45	1-25	yes
x5	46-60	26-49	no
x6	16-30	26-49	yes
x7	46-60	26-49	no

- Let W = {x | Walk(x) = yes}.

$$\underline{A}W = \{x1, x6\},$$

$$\overline{A}W = \{x1, x3, x4, x6\},$$

$$BN_A(W) = \{x3, x4\},$$

$$U - \overline{A}W = \{x2, x5, x7\}.$$

- The decision class, *Walk*, is rough since the boundary region is not empty.

Aproximaciones Inferior y Superior

Ejemplo de DS

U	*Headache*	*Temp.*	*Flu*
U1	Yes	Normal	No
U2	Yes	High	Yes
U3	Yes	Very-high	Yes
U4	No	Normal	No
U5	*No*	*High*	*No*
U6	***No***	***Very-high***	***Yes***
U7	*No*	*High*	*Yes*
U8	***No***	***Very-high***	***No***

$\underline{R}X1$

$\overline{R}X2$

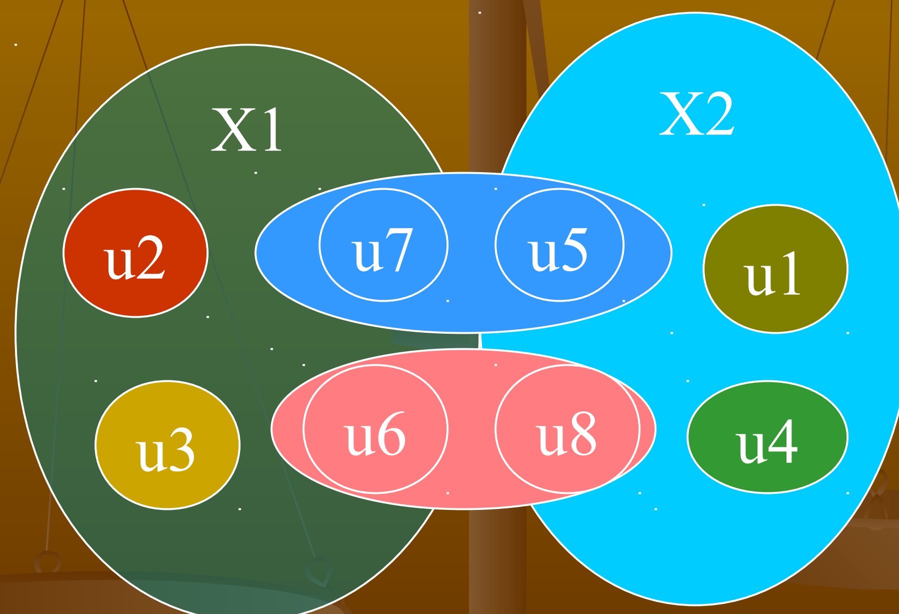
Aproximaciones Inferior y Superior

R = {Headache, Temp.}
U/R = { {u1}, {u2}, {u3}, {u4}, {u5, u7}, {u6, u8}}

X1 = {u | Flu(u) = yes} = *{u2,u3,u6,u7}*
X2 = {u | Flu(u) = no} = *{u1,u4,u5,u8}*

$\underline{R}X1$ = *{u2, u3}*
$\overline{R}X1$ = *{u2, u3, u6, u7,* ***u8, u5****}*

$\underline{R}X2$ = *{u1, u4}*
$\overline{R}X2$ = *{u1, u4, u5, u8,* ***u7, u6****}*



Regiones para más de un atributo de decisión

Dado el sistema de decisión S=(U, A∪D), sean I⊆A y J⊆D, dos relaciones de equivalencias, las cuales inducen sobre U las particiones A*={X1, X2,..., Xn} y D*={Y1, Y2,..., Ym}. Entonces,

POS(D*)= ∪ Bi(Yj), para Yj∈D*,

BND(D*)= ∪ (Bs(Yj)-Bi(Yj)), para Yj∈D*,

NEG(D*)= U - ∪ Bs(Yj), para Yj∈D*,

Estos resultados se pueden expresar de forma general del modo siguiente:

Si F={X1, X2,..., Xn}, entonces POS(F)= POS(X1) ∪ POS(X2) ∪ ... ∪ POS(Xn)

MEDIDAS ASOCIADAS A CONJUNTOS APROXIMADOS

Definición: (Precisión de la aproximación)

$$\alpha_B(X) = \frac{|B_*(X)|}{|B^*(X)|}$$

Donde $|X|$ denota la cardinalidad del conjunto X, y $0 \leq \alpha B(X) \leq 1$. Si $\alpha B(X)=1$, el conjunto X será duro o exacto con respecto a la relación de equivalencia B, mientras que si $\alpha B(X)<1$, el conjunto X es aproximado o vago con respecto a B.

MEDIDAS ASOCIADAS A CONJUNTOS APROXIMADOS (Cont.)

Definición: (Calidad de la aproximación)

$$\gamma_B(X) = \frac{|B_*(X)|}{|X|}$$

Esta medida de calidad representa la frecuencia relativa de los objetos correctamente clasificados por medio de los atributos en B. Además, $0 \leq \alpha B(X) \leq \gamma B(X) \leq 1$.

MEDIDAS ASOCIADAS A CONJUNTOS APROXIMADOS (Cont.)

Medidas unificadas:

$$\sigma_B(X) = \frac{|X \cap POS(X)|}{s1 * |X| + s2 * |POS(X)|}$$

donde s1 y s2 son factores de escala tal que   s1+ s2=1.

MEDIDAS ASOCIADAS A CONJUNTOS APROXIMADOS (Cont.)

Definición: (Calidad de la clasificación)

$$\gamma_B(Y) = \frac{\sum_{i=1}^{n} \left| B_* Yi \right|}{\left| U \right|}$$

Función de pertenencia aproximada

- The rough membership function quantifies the degree of relative overlap between the set *X* and the equivalence class $[x]_B$ to which *x* belongs.

$$\mu_X^B : U \rightarrow [0,1] \qquad \mu_X^B = \frac{|[x]_B \cap X|}{|[x]_B|}$$

- The rough membership function can be interpreted as a frequency-based estimate of $P(x \in X \mid u)$, where *u* is the equivalence class of *IND(B)*.

MEDIDAS ASOCIADAS A CONJUNTOS APROXIMADOS (Cont.)

Definición:(Función de pertenencia aproximada)

$$\mu_X^B(x) = \frac{|X \cap B(x)|}{|B(x)|}$$

Propiedades de esta medida son:

α) μXB(x) = 1 iff x $\in$ Bi(X)

β) μXB(x) = 0 iff x $\in$ U-Bs(X)

c) 0 < μXB(x) < 1 iff x $\in$ BNB(X)

δ) μU-XB(x) = 1-μXB(x) para cualquier x $\in$ U

ε) μX$\cup$YB(x)= max (μXB(x), μYB(x)) para cualquier x $\in$ U

ϕ) μX$\cap$YB(x)= min (μXB(x), μYB(x)) para cualquier x $\in$ U

Aproximaciones (2)

- The formulae for the lower and upper approximations can be generalized to some arbitrary level of precision $\pi \in (0.5, 1]$ by means of the rough membership function

$$\underline{B}_{\pi} X = \{x \mid \mu_X^B(x) \geq \pi\}$$

$$\overline{B}_{\pi} X = \{x \mid \mu_X^B(x) > 1 - \pi\}.$$

- Note: the lower and upper approximations as originally formulated are obtained as a special case with $\pi = 1$.

MEDIDAS ASOCIADAS A CONJUNTOS APROXIMADOS (Cont.)

Definición: (Aspereza del conjunto aproximado)

$$\rho B(X) = 1 - \alpha B(X)$$

MEDIDAS ASOCIADAS A CONJUNTOS APROXIMADOS (Cont.)

Definición: (Entropía áspera)

La entropía áspera (*rough entropy*) de un conjunto aproximado X, denotada por Er(X), se calcula por

$$E_r(X) = -(\rho_B(X))\left[\sum Q_i \log(P_i)\right]$$

para i=1,...,n clases de equivalencias generadas por la relación de equivalencia B.

donde:

El termino ρB(X) denota la aspereza del conjunto aproximado X,

Qi= ni/n, ni es la cantidad de elementos en la clase i y n es la cantidad total de elementos en la combinación de todas las clases,

Pi= 1/ni, ni es la cantidad de elementos en la clase i.

Relación de Subconjunto:

- X esta B-incluido inferiormente en Y, denotado por $X \subseteq *B\ Y$, si y solo si, $Bi(X) \subseteq Bi(Y)$
- X esta B-incluido superiormente en Y, denotado por $X \subseteq *B\ Y$, si y solo si, $Bs(X) \subseteq Bs(Y)$
- X esta B-incluido aproximadamente en Y, denotado por $X \subseteq B\ Y$, si y solo si, $Bi(X) \subseteq Bi(Y)$ y $Bs(X) \subseteq Bs(Y)$, y se dice que X es un B-subconjunto aproximado de Y.

Relación de Igualdad:

- X es B-igual inferiormente a Y, denotado por X =*B Y, si y solo si, Bi(X) = Bi(Y) o (X $\subseteq$*B Y y Y $\subseteq$*B X)
- X es B-igual superiormente a Y, denotado por X =*B Y, si y solo si, Bs(X) = Bs(Y) o (X $\subseteq$*B Y y Y $\subseteq$*B X)
- X es B-igual aproximadamente a Y, denotado por X =B Y, si y solo si, Bi(X) = Bi(Y) y Bs(X) = Bs(Y).

Reducto.

Dado un sistema de información S=(U,A), donde U es el universo y A es el conjunto de atributos, un *reducto* de este es un conjunto mínimo de atributos $B \subseteq A$ tal que $IA = IB$.

Algoritmo QuickReduct

Dado S=(U,A) donde |U|=n, |A|= m.

P1: R={}, A*=A, FIN= false.

P2: Repeat

Seleccionar Ai de A* : f(R∪{Ai}) es

máximo.

Eliminar de A* el Ai seleccionado.

FIN=true If Y(R)=Y(A)

until FIN.

Núcleo de un sistema de decisión

El núcleo de un sistema de decisión (o información) S es el conjunto de atributos B* ⊆A que pertenecen a la intersección de todos los reductos del sistema.

$$Nucleo(S) = \bigcap_{R \in \mathrm{Re}duct(S)} R$$

Un Ejemplo de Reducto & Núcleo

U	*Headache*	*Muscle pain*	*Temp.*	*Flu*
U1	Yes	Yes	Normal	No
U2	Yes	Yes	High	Yes
U3	Yes	Yes	Very-high	Yes
U4	No	Yes	Normal	No
U5	No	No	High	No
U6	No	Yes	Very-high	Yes

Reduct1 = {Muscle-pain,Temp.}

U	*Muscle pain*	*Temp.*	*Flu*
U1,U4	Yes	Normal	No
U2	Yes	High	Yes
U3,U6	Yes	Very-high	Yes
U5	No	High	No

Reduct2 = {Headache, Temp.}

U	*Headache*	*Temp.*	*Flu*
U1	Yes	Norlmal	No
U2	Yes	High	Yes
U3	Yes	Very-high	Yes
U4	No	Normal	No
U5	No	High	No
U6	No	Very-high	Yes

CORE = {Headache,Temp} ∩

{MusclePain,Temp}

= {Temp}

COMPLEJIDAD COMPUTACIONAL DE PROCESOS ASOCIADOS A LOS CONJUNTOS APROXIMADOS

Complejidad computacional de encontrar Bi(X) y Bs(X): $O(lm^2)$,

l es el numero de atributos que describen los objetos

m es la cantidad de objetos en el universo.

Costo computacional de encontrar un reducto: acotado por l^2m^2.

Complejidad en tiempo de encontrar todos los reductos: $O(2^l J)$,

l es la cantidad de atributos,

J es el costo computacional requerido

APLICACIONES DE LOS CONJUNTOS APROXIMADOSAPLICACIONES DE LOS CONJUNTOS APROXIMADOS

- Medicina: diagnostico medico, adquisición de conocimiento, análisis de imágenes medicas, solución de problemas de predicción y clasificación.
- Economía, Finanzas, y Negocios: evaluación de riesgos, análisis de factores de riesgos, predicción de comportamiento, descubrimientos de patrones temporales.
- Estudio del medio ambiente: análisis de datos, predicción de peligros y demandas de recursos naturales.

APLICACIONES DE LOS CONJUNTOS APROXIMADOSAPLICACIONES DE LOS CONJUNTOS APROXIMADOS

- Ingeniería: problemas de control, procesamiento de señales e imágenes, diagnostico de fallas en equipos, análisis de materiales.
- Ciencia de la información: ingeniería de software, recuperación de información, minería de datos, procesamiento del lenguaje natural.
- Ayuda a la decisión: toma de decisiones multi-criterio, planificación inteligente, selección de atributos y ordenamiento, etc..

APLICACIONES DE LOS CONJUNTOS APROXIMADOSAPLICACIONES DE LOS CONJUNTOS APROXIMADOS

- Sociología y Humanidades: análisis de conflictos, problemas de predicción.
- Bioinformática: descubrimiento de patrones, reglas causales, reglas de asociación.
- Química y Farmacia: descubrimiento de estructuras.

Valoración de la aplicabilidad de los conjuntos aproximados en el descubrimiento de conocimiento.

a) Análisis de los atributos a considerar.
- Selección de los atributos.
- Análisis de la dependencia entre atributos.
- Reducción de atributos.
- Calculo de la importancia de un atributo.
- Discretización de atributos.
- Extracción de atributos.

b) Formulación del conocimiento descubierto.
- Descubrimiento de reglas causales.
- Calculo de la certidumbre de las reglas