Ejemplo

Datos de pacientes: La tabla representa un conjunto de datos sobre pacientes que se recobran de una cirugía de la espina dorsal.

Edad – Grupo de edad del paciente

LEMS – índice de movilidad de extremidades inferiores

Walk – Si el paciente camina o no

Object	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	26-29	No
x6	16-30	26-29	Yes
x7	46-60	26-29	No

Ejemplo

Object	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	26-29	No
x6	16-30	26-29	Yes
x7	46-60	26-29	No

Partición Age: {{x1, x2, x6}, {x3,x4}, {x5,x7}}

Partición LEMS : {{x1},{x2},{x3,x4},{x5,x6,x7}}

Partición Walk: {{x1,x4,x6},{x2,x3,x5,x7}}

Note que en este contexto estamos suponiendo:
Clasificación ≡ Partición

Introduccion

- **Clasificación**: Como mecanismo fundamental para expresar conocimiento.
- **Teoretización**: Derivabilidad para obtener nuevo conocimiento a partir del conocimiento que se posee.
- **Reducción**: Procedimiento para eliminar redundancias en la representación del conocimiento.

Teoria de los conjuntos de frontera imprecisa (Rough Set Theory).

Aplicación al análisis de datos.

Dr. Fernando López Irarragorri
Universidad Autónoma de Nuevo León
Facultad de Ingeniería Mecánica y eléctrica
Programa de Posgrado en Ingeniería de sistemas
fernando.lopezrr@uanl.edu.mx.

Agenda

- Filosofia basica de la teoría de los conjuntos de frontera imprecisa .
- Bases de conocimiento.
- Dependencia.
- Reduccion de conocimiento.

Filosofia basica de la teoria de los conjuntos de frontera imprecisa.

- **Universo** como una colección finita de objetos.
- **Conocimiento** como habilidad de discernir (distinguir) objetos de acuerdo con una serie de patrones o atributos.
- **Conceptos** como subconjuntos del universo del discurso.

Filosofia basica de la teoria de los conjuntos de frontera imprecisa.

- **Conocimiento abstracto** como familia de conceptos del universe.
- **Clasificación** como conocimiento abstracto que constituye una partición del universo
- **Categorias** como bloques basicos del conocimiento sobre una situacion particular (conceptos de cierta clasificación).

Filosofia basica de la teoria de los conjuntos de frontera imprecisa.

Ejemplos:

Universo:

U={x1,x2,x3,x4, …,x7}

Conocimiento abstracto:

LEMS_Bajo={{x2},{x3,x4}}

Clasificación:

Age={{x1,x2,x6},{x3,x4},{x5,x7}}

Conceptos:

LEMS_Alto={x1}

Walk_Si={x1,x4,x6}

Categorías:

Age_16-30 = {x1,x2,x6} Age_31-45={x3,x4}

Age_46-60 = {x5,x7}

Object	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	26-29	No
x6	16-30	26-29	Yes
x7	46-60	26-29	No

Filosofia basica de la teoria de los conjuntos de frontera imprecisa.

- **Clasificación** como base sobre la que descansa la armazón teórica de los conjuntos de frontera imprecisa.
- **Aproximación** como herramienta fundamental para el tratamiento de la imprecisión.

Agenda

- Filosofia basica de la teoria de los conjuntos de frontera imprecisa .
- Bases de conocimiento.
- Dependencia.
- Reduccion de conocimiento.

Bases de conocimiento.

- Definición de una base de conocimiento
- Relación de indiscernibilidad.
- Categorias básicas y conocimiento basico.
- Conceptos precisos e imprecisos.
- Aproximación por exceso y aproximación por defecto.
- Región positiva
- Medidas de la aproximación.

Bases de conocimiento.

Sea U el universo, de objetos distinguidos de una cierta realidad, R un conjunto de clasificaciones (información sobre los objetos de U). Se definirá por **base de conocimento** sobre U al par K= (U, R).

$P \subseteq R$ es usualmente denominado conocimiento sobre U.

Bases de conocimiento.

Ejemplo:

Sea *U={Pedro,María,Juana,Octavio,Inés}*,

y *R= {Talla, Peso, Raza, Sexo}*

con

Talla = {{Pedro, Maria}, {Juana, Inés}, {Octavio}},

Peso = {{Pedro, Juana, Inés}, {María}, {Octavio}}.

Raza = {{Pedro}, {María, Juana, Inés,Octavio}},

Sexo = {{Pedro, Octavio}, {María, Juana, Inés}}.

Entonces K=(U,R) es una base de conocimiento sobre U

Bases de conocimiento.

Sea K = (U, R) una base de conocimiento, la **relación de indiscernibilidad** asociada, denotada por IND(K), es aquella clasificación que se obtiene al “interceptar” todas las clasificaciones del conjunto R.

La intercepción de dos clasificaciones A y B está compuesta por las **categorías** que resultan de **interceptar cada categoría de A** con **todas** las **categorías de B** (“Similar al producto de matrices”).

Bases de conocimiento.

Sea U={Pedro,María,Juana,Octavio,Inés},
y R= {Talla, Peso, Raza, Sexo}
con
Talla = {{Pedro, Maria}, {Juana, Inés}, {Octavio}},
Peso = {{Pedro, Juana, Inés}, {María}, {Octavio}}.
Raza = {{Pedro}, {María, Juana, Inés,Octavio}},
Sexo = {{Pedro, Octavio}, {María, Juana, Inés}}.
Entonces K=(U,R) es una base de conocimiento sobre U

Ejemplo:

Sea K=(U,R)

del ejemplo anterior

Talla∩Peso={{Pedro,Maria}∩{Pedro,Juana,Inés},{Pedro, Maria}∩{Maria},{Pedro,Maria}∩{Octavio},{Juana,Inés}∩{Pedro,Juana,Inés}, {Juana,Inés}∩{Maria}, {Juana,Inés}∩ {Octavio}, {Octavio}∩ {Pedro,Juana,Inés}, Octavio}∩{Maria}, Octavio}∩ {Octavio}}

Talla∩Peso={{Pedro},{Maria},∅,{Juana,Inés}, ∅, ∅, ∅, ∅,{Octavio}}

Talla∩Peso = {{Pedro},{Maria}, {Juana,Inés},{Octavio}}

Bases de conocimiento.

Sea *U={Pedro,María,Juana,Octavio,Inés}*,
y *R= {Talla, Peso, Raza, Sexo}*
con

Talla = *{{Pedro, Maria}, {Juana, Inés}, {Octavio}}*,
Peso = *{{Pedro, Juana, Inés}, {María}, {Octavio}}*.
Raza = *{{Pedro}, {María, Juana, Inés,Octavio}}*,
Sexo = *{{Pedro, Octavio}, {María, Juana, Inés}}*.
Entonces K=(U,R) es una base de conocimiento sobre U

Ejemplo:

Sea K=(U,R)

del ejemplo anterior

Talla∩Peso = {{Pedro},{Maria}, {Juana,Inés},{Octavio}}

Talla∩Peso∩Raza= {{Pedro},{Maria}, {Juana,Inés},{Octavio}}

Talla∩Peso∩Raza∩Sexo= Talla∩Peso∩Raza = IND(K)

IND(K)= {{Pedro},{Maria}, {Juana,Inés},{Octavio}}

Bases de conocimiento.

Las categorías de IND(K) son denominadas **categorías básicas**, ya que cualquier otra categoría en K puede ser expresada como una unión finita de tales. Mientras que el conjunto de categorías de IND(K) es denominado **conocimiento básico**.

Bases de conocimiento.

Todo concepto que puede ser expresado como la unión finita de categorías básicas de K es denominado **preciso** en K, si no es posible esto se dice entonces que el concepto es **impreciso** en K.

La imprecisión en la representación de un concepto, en una base de conocimiento K, tiene lugar cuando no se dispone de información suficiente.

Bases de conocimiento.

Sea U={Pedro,María,Juana,Octavio,Inés},
y R= {Talla, Peso, Raza, Sexo}
con
Talla = {{Pedro, Maria}, {Juana, Inés}, {Octavio}},
Peso = {{Pedro, Juana, Inés}, {María}, {Octavio}}.
Raza = {{Pedro}, {María, Juana, Inés,Octavio}},
Sexo = {{Pedro, Octavio}, {María, Juana, Inés}}.
Entonces K=(U,R) es una base de conocimiento sobre U

Ejemplo: Sea K=(U,R)
como en ejemplos anteriores.
IND(K)= {{Pedro},{Maria}, {Juana,Inés}, {Octavio}}

Concepto **Mujer**= {María, Juana, Inés },

nótese que
Mujer = {María}∪{Juana, Inés}, luego **Mujer** es *preciso* en K.

Considerese ahora **Matrimonio** = {Pedro, Juana}
Únicas categorías básicas relacionadas {Pedro} y {Juana, Inés}

Sin embargo **Matrimonio** ≠ {Pedro{Juana, Inés}**,**
luego **Matrimonio** es *impreciso* en K

Conceptos precisos e imprecisos.

Bases de conocimiento.

En el contexto de Rough Sets:

Conjuntos imprecisos se aproximan por un par de conjuntos precisos denominados **aproximación por defecto** y **aproximación por exceso** y de esta manera pueden ser tratados como cualquier concepto preciso.

Bases de conocimiento.

Sea K=(U,R) y X concepto de U

Aproximación por defecto de X: conjunto de todos los elementos de U que pueden ser *clasificados con certeza como elementos de X*

A**proximación por exceso de X:** conjunto de todos aquellos elementos de U que pueden ser *clasificados como posibles elementos de X.*

Bases de conocimiento.

Sea K=(U,R), X ⊆U, P ⊆R

Aproximación por defecto de X según P

$$\underline{P}X = \bigcup_{x_i \in IND(P) \wedge x_i \subseteq X} x_i$$

Aproximación por exceso de X según P

$$\bar{P}X = \bigcup_{x_i \in IND(P) \wedge x_i \cap X \neq \emptyset} x_i$$

Aproximaciones por defecto y por exceso.

Bases de conocimiento.

☛ **Ejemplo:** La aproximación por defecto del concepto **matrimonio** quedaría como:

$\underline{R}X$ = {Pedro}

Mientras que la aproximación por exceso sería:

$\overline{R}X$ = {Pedro, Juana,Inés}

Notese que la aproximación por exceso y la aproximación defecto de **mujer** son idénticas. Esto ocurre para todos los conceptos precisos en K.

O sea, un concepto X solo es preciso o exacto en K si sus aproximaciones coinciden

Bases de conocimiento.

Sea K=(U,R) y X concepto impreciso en K

Sean $\underline{R}X$ y $\overline{R}X$ las aproximaciones por defecto y por exceso de X en K

La región positiva de X se define como:
POS_R(X)=$\bigcup_{x_i \in \overline{R}X} x_i$

La frontera imprecisa de X se define como:
B_R(X)= $\overline{R}X \setminus \underline{R}X$

Le región negativa de X se define como:
NEG_R(X)=U \ $\overline{R}X$

Medidas de la aproximacion.

Bases de conocimiento.

Sea *U={Pedro,María,Juana,Octavio,Inés}*,
y *R= {Talla, Peso, Raza, Sexo}*
con
Talla = *{{Pedro, Maria}, {Juana, Inés}, {Octavio}}*,
Peso = *{{Pedro, Juana, Inés}, {María}, {Octavio}}*.
Raza = *{{Pedro}, {María, Juana, Inés,Octavio}}*,
Sexo = *{{Pedro, Octavio}, {María, Juana, Inés}}*.
Entonces K=(U,R) es una base de conocimiento sobre U

Ejemplo:
Del ejemplo anterior se tiene:

POS_R(**Matrimonio**)={Pedro}

B_R(**Matrimonio**) = {Juana,Inés}

NEG_R(Matrimonio)={Maria, Octavio}

Medidas de la aproximacion.

Bases de conocimiento.

Sea K=(U,R) y X concepto impreciso en K

Se define como calidad de la aproximación de X en K

$$\mu_R(X) = \frac{card(\underline{R}X)}{card(\overline{R}X)}$$

Bases de conocimiento.

La aproximación de conceptos puede extenderse de manera natural a las clasificaciones, así

Si F = $\{X_1, X_2,...,X_n\}$ es una clasificación. Entonces se definen las aproximaciones por defecto y por exceso de F como sigue:

$$\underline{R}F = \{\underline{R}X_1, ..., \underline{R}X_n\}$$

$$\overline{R}F = \{\overline{R}X_1, ..., \overline{R}X_n\}$$

La union de los $\overline{R}X_i$ se denomina region positiva

Bases de conocimiento.

Para medir la “imprecisión” en la representación de una clasificación F se definirán dos medidas:

la “exactitud” de la aproximación:

$$\alpha_R(F) = \frac{\sum Card\,\underline{R}X_i}{\sum Card\,\overline{R}X_i}$$

y la “calidad” de la aproximación:

$$\gamma_R(F) = \frac{\sum Card\,\underline{R}X_i}{Card\,U}$$

Medidas de la aproximacion.

Bases de conocimiento.

- La **exactitud de la aproximación** expresa el porcentaje de las posibles decisiones correctas cuando se clasifican objetos en clases de F utilizando conocimiento R.
- La **calidad de la aproximación** expresa el porcentaje de objetos que pueden ser correctamente clasificados en clases de F utilizando conocimiento R.

Bases de conocimiento.

☛**Ejemplo:** Considere la clasificación **ocupación = {{Pedro},{Maria, Juana},{Inés}, {Octavio}}** que es externa a K. Se tienen las siguientes aproximaciones:

$$\underline{R}O = \{\{Pedro\},\{Maria\},\{\varnothing\},\{Octavio\}\}$$

$$\overline{R}O = \{\{Pedro\},\{Maria\},\{Juana, Ines\},\{Octavio\}\}$$

entonces las medidas de la imprecisión quedarían como:

$$\alpha_R(F) = \frac{1+1+0+1}{1+1+2+1} = 3/5 \qquad \gamma_R(F) = \frac{1+1+1+0+1}{5} = 3/5$$

Sumario

- Filosofia basica de la teoria de los conjuntos de frontera imprecisa .
- Bases de conocimiento.
- Dependencia.
- Reduccion de conocimiento.

Dependencia.

- Concepto de dependencia, equivalencia e independencia.
- Concepto de dependencia parcial.

Dependencia.

Sea K=(U,R) y P,Q $\subseteq$R entonces

Q depende o es derivable de P sii

IND(P) $\subseteq$ IND(Q),

Cada categoría básica de P está contenida en una categoría básica de Q

también se denota por

P$\Rightarrow Q$

Dependencia.

Ejemplo:
Sea K=(U,R) y P,Q $\subseteq$R

Sean dados los conocimientos

P = {{1,5}, {2,8}, {3},{4},{6},{7}}
Q = {{1,5},{2,7,8},{3,4,6}}

Por simple inspección se observa que IND(P) $\Longrightarrow$ IND(Q), luego se concluye que el conocimiento Q depende del conocimiento P, o sea P $\Rightarrow$ Q.

Dependencia.

La derivación (dependencia) del conocimiento tambien puede ser expresada de manera parcial. Para tal efecto se usa la noción de región positiva del conocimiento.

Se dice que Q depende en grado k de P, simbolicamente $P \Rightarrow_k Q$, si y solo si:

$$k = \gamma_P(Q) = \frac{Card POS_P(Q)}{Card U}$$

Dependencia parcial

Dependencia.

☛**Ejemplo:** Considerando P como en el ejemplo anterior y a Q = {{1},{2,7},{3,6},{4},{5,8}}, se obtiene:

$$\overline{P}Q = \{\{\emptyset\},\{7\},\{3,6\},\{4\},\{\emptyset\}\}$$

luego,

$$POS_P(Q) = \{3,4,6,7\}$$

de aqui que

$$k = 4/8$$

Sumario

- Filosofia basica de la teoria de los conjuntos de frontera imprecisa .
- Bases de conocimiento.
- Dependencia.
- Reduccion de conocimiento.

Reduccion de conocimiento

- Dispensabilidad e indispensabilidad.
- Reductos (Reducts).
- Nucleo (Core).

Reduccion de conocimiento

Sea R una familia de relaciones de equivalencia y sea $P \in R$, se dice que P es **dispensable** en R si IND(R) = IND(R-{P}). En otro caso se dice que P es indispensable.

La familia R se dice **independiente** si cada elemento P de la misma es indispensable, de lo contrario se dice que R es **dependiente**.

Dispensabilidad e indispensabilidad

Reduccion de conocimiento

- **Ejemplo:** Supongase que se tiene una familia R={P,Q, F} de clasificaciones de una base de conocimiento K. Donde:

P = {{1,4,5},{2,8},{3},{6,7}}

Q= {{1,3,5},{6},{2,4,7,8}}

F = {{1,5},{6},{2,7,8},{3,4}}

luego

IND(R) = {{1,5},{2,8},{3,},{4},{6},{7}}

IND(R-{P}) = {{1,5},{2,7,8},{3},{4},{6}} $\neq$ IND(R)

por lo tanto se concluye que P es indispensable.

De igual manera se tiene que Q y F son dispensables. Esto implica que R es dependiente.

Reduccion de conocimiento

El conocimiento Q $\subseteq$ P, se denomina **reducto** de P si:
Q es independiente
y IND(Q)=IND(P).

Obviamente P puede tener varios reductos.

☛**Ejemplo:** En el ejemplo anterior se tienen dos reductos de R: {P,Q} y {P,F}.

Reduccion de conocimiento

Sea R una familia de clasificaciones. El conjunto de todas las clasificaciones indispensables se denomina **nucleo** de R se denota por CORE(R).

Una definición equivalente de nucleo puede ser dada como: la intersección de todos los reductos de R.

☛**Ejemplo:** Tomando R como se supuso en el ejemplo anterior se tiene que el único elemento indispensable en R es P, luego
CORE(R) = {P}

El núcleo de una familia de clasificaciones puede ser vacio.

Nucleo

Ejemplo

Suponga que debe comprar un carro, y que se le presentan las siguientes alternativas:

Auto	Precio	Consumo	Tamaño	Max. Vel.	Accel.
1	bajo	medio	grande	baja	buena
2	medio	medio	comp.	alta	pobre
3	alto	medio	grande	baja	buena
4	medio	medio	grande	alta	exc.
5	bajo	alto	grande	baja	buena

Ejemplo

Precio = {{1,5},{2,4}, {3}},

Consumo = {{1,2,3,4},{5}},

Tamaño = {{2},{1,3,4,5}},

Maxima Velocidad = {{1,3,5},{2,4}},

Aceleracion = {{1,3,5},{2},{4}}.

IND(K) = {{1},{2},{3},{4},{5}}

Ejemplo

Analicemos dependencia entre:

Precio (P) y Máxima Velocidad (V), o sea deseamos

Saber si $P \Longrightarrow_k$ Q o viceversa

Por simple inspección

$P \Longrightarrow$ Q

Analicemos ahora $Q \Longrightarrow_k P$

$POS_Q(P) = \{2,4\}$, luego k=2/5

Precio = {{1,5},{2,4}, {3}},
Consumo = {{1,2,3,4},{5}},
Tamaño = {{2},{1,3,4,5}},
Maxima Velocidad = {{1,3,5},{2,4}},
Aceleracion = {{1,3,5},{2},{4}}.

IND(K) = {{1},{2},{3},{4},{5}}

Ejemplo

Redundancia:
IND({Precio,Consumo, Tamaño}) = IND(K)

IND({Precio,Consumo}) = {{1},{2,4},{3},{5}}
IND({Precio,Consumo}) $\neq$ IND(K)
Tamaño es indispensable

IND({Precio,Tamaño})= {{1,5},{2},{3},{4}}
IND({Precio, Tamaño}) $\neq$ IND(K)
Consumo es indispensable

IND({Tamaño, Consumo}) = {{1,3},{2},{4},{5}}
IND({Tamaño,Consumo}) $\neq$ IND(K)
Precio es indispensable

Precio = {{1,5},{2,4}, {3}},
Consumo = {{1,2,3,4},{5}},
Tamaño = {{2},{1,3,4,5}},
Maxima Velocidad = {{1,3,5},{2,4}},
Aceleracion = {{1,3,5},{2},{4}}.

IND(K) = {{1},{2},{3},{4},{5}}

{Precio,Consumo, Tamaño}
Es un reducto en K

Ejecutar Jupyter notebook en línea

https://hub.mybinder.org/user/binder-examples-r-vwzq7zls/notebooks/index.ipynb#