Normalización
Clase Práctica
SPI y SPDF

Andrea Manna

Departamento de Computación - Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Base de Datos
1er. Cuatrimestre 2017

Esquema General

Dos Características de una Buena Descomposición

- Sin Pérdida de Información
- Preservación de Dependencias Funcionales

Definición

Si R es un esquema de relación descompuesto en los esquemas $R_1, R_2, ..., R_k$ y F es un conjunto de dependencias, decimos que la descomposición es sin pérdida de información (SPI) con respecto a F, si para toda relación r para R que satisfaga F:

$$r = \pi_{R_1}(r) \bowtie \pi_{R_2}(r) \bowtie \ldots \bowtie \pi_{R_k}(r)$$

Es decir, no debe perderse información en el proceso de descomposición, de manera tal que r es la junta natural de sus proyecciones sobre los R_i

Estrategias para comprobar SPI

Descomposiciones de dos esquemas

Descomposición binaria

Descomposiciones en más de dos esquemas

Algoritmo del Tableau

Teorema de la Descomposición Binaria

La descomposición ρ de R, ρ = (R_1, R_2) es SPI respecto a un conjunto de dependencias funcionales F sí y sólo sí:

F^+ contiene la DF: $R_1 \cap R_2 \rightarrow (R_1 - R_2)$

o

F^+ contiene la DF: $R_1 \cap R_2 \rightarrow (R_2 - R_1)$

Ejemplo

Sea R = (A, B, C) y F = $\{A \rightarrow B\}$.

Pregunta: La descomposición de R en $R_1 = (A, B)$ y $R_2 = (A, C)$ es SPI?

Resp: Sí. $AB \cap AC$ = A, AB - AC = B, y $A \rightarrow B$ está en F^+.

Para pensar:

La descomposición de R en $R_1 = (A, B)$ y $R_2 = (B, C)$ es SPI?

Ejemplo

No todas las descomposiciones son sin pérdida, dado que existen proyecciones cuya reunión no dan exactamente la relación original. Ejemplo:

N_Empleado	Funcion	N_Proyecto
Lopez	diseñador	Nueva España
Lopez	programador	Emprendedor
Lopez	diseñador	Emprendedor
Perez	diseñador	Emprendedor

Se puede descomponer en dos tablas:

N_Empleado	Funcion
Lopez	diseñador
Lopez	programador
Perez	diseñador

Funcion	N_Proyecto
diseñador	Nueva España
programador	Emprendedor
diseñador	Emprendedor

Ejemplo

Sin embargo, cuando se reunen las dos tablas, se obtienen tuplas adicionales que no estaban en la original. Estas tuplas se llaman espúreas creadas por los procesos de proyección y reunión. Ya que sin la tabla original, no hay forma de identificar cuáles tuplas son genuinas y cuáles espúreas, se puede perder información (aún cuando se tienen más tuplas) si se sustituyen las proyecciones para la relación original:

N_Empleado	Funcion	N_Proyecto
Lopez	diseñador	Nueva España
Lopez	programador	Emprendedor
Lopez	diseñador	Emprendedor
Perez	diseñador	Nueva España
Perez	diseñador	Emprendedor

Ejemplo

¿La descomposición de R en R_1=(A, D, E, F) y R_2= (B, C, D) es SPI respecto de FD1?

Recordar...

R=(A, B, C, D, E, F) y FD1= { A $\rightarrow$ BD, B $\rightarrow$ CD, AC $\rightarrow$ E}

$R_1 \bigcap R_2$ = {D}
$R_1 - R_2$= {A, E, F}
$R_2 - R_1$= {B, C}
Tenemos que hacer D^+ pero D^+=D y por lo tanto no se cumple que D$\rightarrow${A, E, F} ni que D$\rightarrow${B, C}
Es decir, ninguna de esas dos dependencias funcionales está en $FD1^+$
Por lo tanto la descomposición no es SPI

Definición de Tableau

Dado $R = (A_1, \ldots, A_n)$, un tableau T para una descomposición $\rho = (R_1,\ldots,R_k)$ de R se define de la siguiente forma:

1. T tiene n columnas, una para cada atributo de R
2. T tiene k filas, una para cada esquema de ρ
3. Dadas la fila i y la columna j (esquema R_i y atributo A_j), el contenido del tableau será:

$$a_j \text{ si } A_j \in R_i$$
$$\text{o}$$
$$b_{ij} \text{ si } A_j \notin R_i$$

Los a_j se denominan símbolos distinguidos, y los b_{ij} no distinguidos.

Algoritmo del Tableau

INPUT: Un esquema de relación R, un conjunto de dependencias funcionales F, y una descomposición ρ. OUTPUT: Una decisión de si ρ es SPI.

Construir el Tableau T
mientras haya cambios sobre T
    para cada df $X \rightarrow Y \in F$
        buscar filas que coincidan en todos los símbolos de X
            Si se encontrasen dos filas, igualar los simbolos para los atributos de Y. Cuando se igualan 2 símbolos, si alguno de ellos es a_j, asignarle al otro a_j. Si ellos son b_{ij} y b_{lj}, asignarle a ambos b_{ij} o b_{lj}.
        Si hay una fila con todos símbolos distinguidos, retornar Sí
end (mientras)
Retornar No

Verificación SPI - Ejercicio 1

Sea R = (A,B,C,D,E) y F = $\{A \rightarrow B, D \rightarrow C\}$.
Decidir si la descomposición en
R_1 = (A,B,C), R_2 = (C,D,E) y R_3 = (A,D,E) es SPI.

Verificación SPI - Ejercicio 1 - Tableau Inicial

R = (A,B,C,D,E), F = $\{A \to B, D \to C\}$, R_1 = (A,B,C), R_2 = (C,D,E) y R_3 = (A,D,E)

	A	**B**	**C**	**D**	**E**
ABC	a_1	a_2	a_3	b_{14}	b_{15}
CDE	b_{21}	b_{22}	a_3	a_4	a_5
ADE	a_1	b_{32}	b_{33}	a_4	a_5

Verificación SPI - Ejercicio 1 - Tableau Intermedio

R = (A,B,C,D,E), F = $\{A \rightarrow B, D \rightarrow C\}$, R_1 = (A,B,C), R_2 = (C,D,E) y R_3 = (A,D,E)

$$A \rightarrow B$$

	A	**B**	**C**	**D**	**E**
ABC	a_1	a_2	a_3	b_{14}	b_{15}
CDE	b_{21}	b_{22}	a_3	a_4	a_5
ADE	a_1	b_{32}	b_{33}	a_4	a_5

Verificación SPI - Ejercicio 1 - Tableau Intermedio

R = (A,B,C,D,E), F = $\{A \to B, D \to C\}$, R_1 = (A,B,C), R_2 = (C,D,E) y R_3 = (A,D,E)

$A \to B$

	A	B	C	D	E
ABC	a_1	a_2	a_3	b_{14}	b_{15}
CDE	b_{21}	b_{22}	a_3	a_4	a_5
ADE	a_1	a_2	b_{33}	a_4	a_5

Verificación SPI - Ejercicio 1 - Tableau Intermedio

R = (A,B,C,D,E), F = $\{A \to B, D \to C\}$, R_1 = (A,B,C), R_2 = (C,D,E) y R_3 = (A,D,E)

$D \to C$

	A	B	C	D	E
ABC	a_1	a_2	a_3	b_{14}	b_{15}
CDE	b_{21}	b_{22}	a_3	a_4	a_5
ADE	a_1	a_2	b_{33}	a_4	a_5

Verificación SPI - Ejercicio 1 - Tableau Final

R = (A,B,C,D,E), F = $\{A \to B, D \to C\}$, R_1 = (A,B,C), R_2 = (C,D,E) y R_3 = (A,D,E)

$$D \to C$$

	A	**B**	**C**	**D**	**E**
ABC	a_1	a_2	a_3	b_{14}	b_{15}
CDE	b_{21}	b_{22}	a_3	a_4	a_5
ADE	a_1	a_2	a_3	a_4	a_5

Verificación SPI - Ejercicio 1 - Tableau Final

R = (A,B,C,D,E), F = $\{A \rightarrow B, D \rightarrow C\}$, R_1 = (A,B,C), R_2 = (C,D,E) y R_3 = (A,D,E)

	A	B	C	D	E
ABC	a_1	a_2	a_3	b_{14}	b_{15}
CDE	b_{21}	b_{22}	a_3	a_4	a_5
ADE	a_1	a_2	a_3	a_4	a_5

Como hay una fila con todos símbolos distinguidos, ρ es SPI.

Verificación SPI - Ejercicio 2

Sea R = (A,B,C,D,E,F,G,H,I),
F = $\{A \to B, CD \to F, H \to AD, I \to C, D \to H\}$.
Decidir si la descomposición $\rho = \{ABD, DEF, FGC, CHI\}$ es SPI.

Verificación SPI - Ejercicio 2 - Tableau Inicial

R = (A,B,C,D,E,F,G,H,I),
F = $\{A \rightarrow B, CD \rightarrow F, H \rightarrow AD, I \rightarrow C, D \rightarrow H\}$,
$\rho = \{ABD, DEF, FGC, CHI\}$

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
ABD	a_1	a_2	b_{13}	a_4	b_{15}	b_{16}	b_{17}	b_{18}	b_{19}
DEF	b_{21}	b_{22}	b_{23}	a_4	a_5	a_6	b_{27}	b_{28}	b_{29}
FGC	b_{31}	b_{32}	a_3	b_{34}	b_{35}	a_6	a_7	b_{38}	b_{39}
CHI	b_{41}	b_{42}	a_3	b_{44}	b_{45}	b_{46}	b_{47}	a_8	a_9

Verificación SPI - Ejercicio 2 - Tableau Intermedio

R = (A,B,C,D,E,F,G,H,I),
F = $\{A \rightarrow B, CD \rightarrow F, H \rightarrow AD, I \rightarrow C, D \rightarrow H\}$,
$\rho = \{ABD, DEF, FGC, CHI\}$

$$D \rightarrow H$$

	A	B	C	D	E	F	G	H	I
ABD	a_1	a_2	b_{13}	a_4	b_{15}	b_{16}	b_{17}	b_{18}	b_{19}
DEF	b_{21}	b_{22}	b_{23}	a_4	a_5	a_6	b_{27}	b_{28}	b_{29}
FGC	b_{31}	b_{32}	a_3	b_{34}	b_{35}	a_6	a_7	b_{38}	b_{39}
CHI	b_{41}	b_{42}	a_3	b_{44}	b_{45}	b_{46}	b_{47}	a_8	a_9

Verificación SPI - Ejercicio 2 - Tableau Intermedio

R = (A,B,C,D,E,F,G,H,I),
F = $\{A \to B, CD \to F, H \to AD, I \to C, D \to H\}$,
$\rho = \{ABD, DEF, FGC, CHI\}$

$D \to H$

	A	B	C	D	E	F	G	H	I
ABD	a_1	a_2	b_{13}	a_4	b_{15}	b_{16}	b_{17}	b_{18}	b_{19}
DEF	b_{21}	b_{22}	b_{23}	a_4	a_5	a_6	b_{27}	b_{18}	b_{29}
FGC	b_{31}	b_{32}	a_3	b_{34}	b_{35}	a_6	a_7	b_{38}	b_{39}
CHI	b_{41}	b_{42}	a_3	b_{44}	b_{45}	b_{46}	b_{47}	a_8	a_9

Verificación SPI - Ejercicio 2 - Tableau Intermedio

R = (A,B,C,D,E,F,G,H,I),
F = $\{A \to B, CD \to F, H \to AD, I \to C, D \to H\}$,
$\rho = \{ABD, DEF, FGC, CHI\}$

$H \to AD$

	A	B	C	D	E	F	G	H	I
ABD	a_1	a_2	b_{13}	a_4	b_{15}	b_{16}	b_{17}	b_{18}	b_{19}
DEF	b_{21}	b_{22}	b_{23}	a_4	a_5	a_6	b_{27}	b_{18}	b_{29}
FGC	b_{31}	b_{32}	a_3	b_{34}	b_{35}	a_6	a_7	b_{38}	b_{39}
CHI	b_{41}	b_{42}	a_3	b_{44}	b_{45}	b_{46}	b_{47}	a_8	a_9

Verificación SPI - Ejercicio 2 - Tableau Intermedio

R = (A,B,C,D,E,F,G,H,I),
F = $\{A \to B, CD \to F, H \to AD, I \to C, D \to H\}$,
$\rho = \{ABD, DEF, FGC, CHI\}$

$$H \to AD$$

	A	B	C	D	E	F	G	H	I
ABD	a_1	a_2	b_{13}	a_4	b_{15}	b_{16}	b_{17}	b_{18}	b_{19}
DEF	a_1	b_{22}	b_{23}	a_4	a_5	a_6	b_{27}	b_{18}	b_{29}
FGC	b_{31}	b_{32}	a_3	b_{34}	b_{35}	a_6	a_7	b_{38}	b_{39}
CHI	b_{41}	b_{42}	a_3	b_{44}	b_{45}	b_{46}	b_{47}	a_8	a_9

Verificación SPI - Ejercicio 2 - Tableau Intermedio

R = (A,B,C,D,E,F,G,H,I),
F = $\{A \rightarrow B, CD \rightarrow F, H \rightarrow AD, I \rightarrow C, D \rightarrow H\}$,
$\rho = \{ABD, DEF, FGC, CHI\}$

$A \rightarrow B$

	A	B	C	D	E	F	G	H	I
ABD	a_1	a_2	b_{13}	a_4	b_{15}	b_{16}	b_{17}	b_{18}	b_{19}
DEF	a_1	b_{22}	b_{23}	a_4	a_5	a_6	b_{27}	b_{18}	b_{29}
FGC	b_{31}	b_{32}	a_3	b_{34}	b_{35}	a_6	a_7	b_{38}	b_{39}
CHI	b_{41}	b_{42}	a_3	b_{44}	b_{45}	b_{46}	b_{47}	a_8	a_9

Verificación SPI - Ejercicio 2 - Tableau Intermedio

R = (A,B,C,D,E,F,G,H,I),
F = $\{A \rightarrow B, CD \rightarrow F, H \rightarrow AD, I \rightarrow C, D \rightarrow H\}$,
$\rho = \{ABD, DEF, FGC, CHI\}$

$$A \rightarrow B$$

	A	B	C	D	E	F	G	H	I
ABD	a_1	a_2	b_{13}	a_4	b_{15}	b_{16}	b_{17}	b_{18}	b_{19}
DEF	a_1	a_2	b_{23}	a_4	a_5	a_6	b_{27}	b_{18}	b_{29}
FGC	b_{31}	b_{32}	a_3	b_{34}	b_{35}	a_6	a_7	b_{38}	b_{39}
CHI	b_{41}	b_{42}	a_3	b_{44}	b_{45}	b_{46}	b_{47}	a_8	a_9

Verificación SPI - Ejercicio 2 - Tableau Final

R = (A,B,C,D,E,F,G,H,I),
F = $\{A \to B, CD \to F, H \to AD, I \to C, D \to H\}$,
$\rho = \{ABD, DEF, FGC, CHI\}$

	A	B	C	D	E	F	G	H	I
ABD	a_1	a_2	b_{13}	a_4	b_{15}	b_{16}	b_{17}	b_{18}	b_{19}
DEF	a_1	a_2	b_{23}	a_4	a_5	a_6	b_{27}	b_{18}	b_{29}
FGC	b_{31}	b_{32}	a_3	b_{34}	b_{35}	a_6	a_7	b_{38}	b_{39}
CHI	b_{41}	b_{42}	a_3	b_{44}	b_{45}	b_{46}	b_{47}	a_8	a_9

Como no hay ninguna fila con todos símbolos distinguidos, y aunque sigamos iterando nuevamente por todas las dependencias, ninguna alterará el tableau, ρ NO es SPI.

Preservación de Dependencias Funcionales

Dados un esquema de relación R, una descomposición $\rho = (R_1, \ldots, R_k)$, y un conjunto F de dependencias funcionales.

$\pi_Z(F)$: proyección de F sobre un conjunto de atributos Z

Conjunto de dependencias $X \rightarrow Y$ en F^+ tal que $XY \subseteq Z$

Testeo (orden exponencial)

La descomposición ρ preserva F si $F^+ = (\bigcup_{i=1}^{k} \pi_{R_i}(F))^+$

Es decir, la descomposición ρ preserva el conjunto de dependencias F si la unión de todas las dependencias en $\pi_{R_i}(F)$ implica lógicamente a todas las dependencias en F

Testeo Polinomial de Preservación de Dependencias Funcionales

Dados un esquema de relación R, una descomposición $\rho = (R_1, \ldots, R_k)$, y un conjunto F de dependencias funcionales.

Para toda dependencia funcional $X \to Y \in F$:

 Verificar que se preserva $X \to Y$:

  Z = X

  while Z cambia

   for i = 1 to k do

    /* clausura con respecto a F */

    Z = $Z \cup ((Z \cap R_i)^+ \cap R_i)$

  Si $Y \not\subseteq Z$ retornar No

Retornar Sí

Preservación de Dependencias Funcionales: Ejercicio

Sean $R = (A, B, C, D, E)$ y
$F = \{AB \rightarrow C, A \rightarrow D, D \rightarrow E, E \rightarrow C\}$.

Decidir si la descomposición $\rho = \{AD, DE, ECB\}$ es sin pérdida de dependencias funcionales (SPDF).

Preservación de Dependencias Funcionales: Resolución Ejercicio

$R = (A, B, C, D, E)$
$F = \{AB \rightarrow C, A \rightarrow D, D \rightarrow E, E \rightarrow C\}$
$\rho = \{AD, DE, ECB\}$

Estrategia de Resolución:
Las dependencias $A \rightarrow D, D \rightarrow E, E \rightarrow C$ se preservan trivialmente (por qué?), y no es necesario aplicarles el algoritmo.
Le aplicaremos el algoritmo a la dependencia $AB \rightarrow C$ para ver si se preserva.

Preservación de Dependencias Funcionales: Resolución Ejercicio

$R = (A, B, C, D, E)$
$F = \{AB \to C, A \to D, D \to E, E \to C\}$
$\rho = \{AD, DE, ECB\}$
Queremos verificar que se preserva $AB \to C$

Verificar que se preserva $X \to Y$:

Z = X
while Z cambia
for i = 1 to k do
$Z = Z \cup ((Z \cap R_i)^+ \cap R_i)$
Si $Y \not\subseteq Z$ retornar No

Z = AB

Preservación de Dependencias Funcionales: Resolución Ejercicio

$R = (A, B, C, D, E)$
$F = \{AB \rightarrow C, A \rightarrow D, D \rightarrow E, E \rightarrow C\}$
$\rho = \{AD, DE, ECB\}$
Queremos verificar que se preserva $AB \rightarrow C$

Verificar que se preserva $X \rightarrow Y$:

    Z = X
    while Z cambia
        for i = 1 to k do
            $Z = Z \cup ((Z \cap R_i)^+ \cap R_i)$
    Si $Y \not\subseteq Z$ retornar No

$$\begin{aligned} Z = Z \cup ((Z \cap R_1)^+ \cap R_1) &= \{A, B\} \cup ((\{A, B\} \cap \{A, D\})^+ \cap \{A, D\}) \\ &= \{A, B\} \cup ((A)^+ \cap \{A, D\}) \\ &= \{A, B\} \cup (\{A, D, E, C\} \cap \{A, D\}) \\ &= \{A, B, D\} \end{aligned}$$

C no está incluido en Z; seguimos...

Preservación de Dependencias Funcionales: Resolución Ejercicio

$R = (A, B, C, D, E)$
$F = \{AB \to C, A \to D, D \to E, E \to C\}$
$\rho = \{AD, DE, ECB\}$
Queremos verificar que se preserva $AB \to C$

Verificar que se preserva $X \to Y$:

Z = X
while Z cambia
  for i = 1 to k do
    $Z = Z \cup ((Z \cap R_i)^+ \cap R_i)$
Si $Y \not\subseteq Z$ retornar No

$$\begin{aligned} Z = Z \cup ((Z \cap R_2)^+ \cap R_2) &= \{A, B, D\} \cup ((\{A, B, D\} \cap \{D, E\})^+ \cap \{D, E\}) \\ &= \{A, B, D\} \cup ((D)^+ \cap \{D, E\}) \\ &= \{A, B, D\} \cup (\{D, E, C\} \cap \{D, E\}) \\ &= \{A, B, D, E\} \end{aligned}$$

C no está incluido en Z; seguimos...

Preservación de Dependencias Funcionales: Resolución Ejercicio

$R = (A, B, C, D, E)$
$F = \{AB \to C, A \to D, D \to E, E \to C\}$
$\rho = \{AD, DE, ECB\}$
Queremos verificar que se preserva $AB \to C$

Verificar que se preserva $X \to Y$:

Z = X
while Z cambia
  for i = 1 to k do
    Z = $Z \cup ((Z \cap R_i)^+ \cap R_i)$

Si $Y \not\subseteq Z$ retornar No

$$Z = Z \cup ((Z \cap R_3)^+ \cap R_3) =$$
$$\{A, B, D, E\} \cup ((\{A, B, D, E\} \cap \{E, C, B\})^+ \cap \{E, C, B\})$$
$$= \{A, B, D, E\} \cup ((EB)^+ \cap \{E, C, B\})$$
$$= \{A, B, D, E\} \cup (\{E, B, C\} \cap \{E, C, B\})$$
$$= \{A, B, D, E, C\}$$

Ahora sí C está incluido en Z: la dependencia $AB \to C$ se preserva

Claves

Algoritmo para encontrar todas las claves

Encontrar todas las claves

1. Obtener el conjunto S de atributos que no figuran en un lado derecho de una DF
2. Verificar si ese conjunto es superclave. Si lo es, es clave UNICA!
3. Si no lo era, agregar paulatinamente a S todas las combinaciones posibles de subconjuntos de R-S (todos los de cardinalidad 1, luego de los de 2, etc) (llamémoslo S') y verificar si cada uno de esos conjuntos es superclave. En este paso se deben obviar todos aquellos S' que contienen una superclave ya calculada, ya que no van a ser minimales.

   Todos los conjuntos de atributos obtenidos que determinan a todo R son las claves.

Ejercicio 1

Recordar...

R=(A, B, C, D, E, F) y FD1= { A $\rightarrow$ BD, B $\rightarrow$ CD, AC $\rightarrow$ E}

Tomamos S=AF y calculamos AF^{+}_{FD1}
AF^{+}_{FD1} = AFBDCE
Como es igual a todo R, es clave UNICA! y el algoritmo termina

Ejercicio 2

Recordar...

R=(A, B, C, D, E, F) y FD2= { A $\to$ BD, B $\to$ CD, AC $\to$ E, C $\to$ A}

Tomamos S=F y calculamos F^{+}_{FD2}
F^{+}_{FD2} = F. Como no es todo R, no es clave
Verificamos con un atributo adicional:

FA^{+}_{FD2} = FABDCE. Es igual a R, es clave, pero seguimos con otro atributo

FB^{+}_{FD2} = FBCDAE. Es igual a R, es clave, pero seguimos con otro atributo

FC^{+}_{FD2} = FCABDE. Es igual a R, es clave, pero seguimos con otro atributo

FD^{+}_{FD2} = FD. No es igual a R, entonces no es clave

FE^{+}_{FD2} = FE. No es igual a R, entonces no es clave

No tiene sentido agregarle A, B ó C a FD y FE. Por qué?

Falta FDE^{+}_{FD2} = FDE. No es igual a R, entonces no es clave. Idem con FED

Ejercicio 3

IMPORTANTE: Es un error pensar que la clave minimal es la que tiene menos atributos

Recordar... Una clave es minimal si ningún subconjunto de ella es clave. No está relacionado con el tamaño de las otras claves

Supongamos lo siguiente: R=(A, B, C, D) y FD= { AB $\rightarrow$ CD, C $\rightarrow$ AB}
Tomamos S=vacío (todos los atributos están en la parte derecha de alguna FD)
A^{+}_{FD} = A. A no es clave
B^{+}_{FD} = B. B no es clave
C^{+}_{FD} = CABD. Es igual a R, es clave, pero seguimos...
D^{+}_{FD} = D. D no es clave. Hay que seguir buscando
AB^{+}_{FD} = ABCD. AB es clave
AD^{+}_{FD} = AD. AD no es clave. Idem con BD y ahora sí no se puede seguir más (Por que?)

Bibliografía

Referencia

- Elmasri/Navathe - Fundamentals of Database Systems, 7th Ed., Pearson, 2016 (Parte 6)
- Ullman - Principles of Database and Knowledge Base Systems, Computer Science Press, 1988 (Capítulo 7)

Bibliografía

No se vayan que ahora viene la segunda parte de la clase!!!