

Adatbáziskezelés Hűségesség

Csima Judit

BME, VIK,
Számítástudományi és Információelméleti Tanszék

2018. október 26. és november 7.

Felbontások

Cél: Adott (R, F) sémából anomáliát nem tartalmazó olyan felbontás előállítása, amiből ugyanaz az információ nyerhető, mint az eredetiből.

Definíció

$\rho = (R_1, \dots, R_k)$ az (R, F) séma felbontása, ha $R_i \subseteq R$ és $\bigcup_{i=1}^k R_i = R$.
Ha r egy (R, F) sémára illeszkedő reláció, akkor legyen $r_i = \pi_{R_i}(r)$ és

$$m_\rho(r) := r_1 \bowtie r_2 \bowtie \dots \bowtie r_k$$

(Megj.: \bowtie asszociatív, így nem kell a zárójelezéssel vesződni)

Kérdés: mikor nyerhető vissza az infó a felbontásból? Mi általában r és $m_\rho(r)$ viszonya?

Példa

Legyen (R, F) a következő: $R(A, B, C)$, $F = \{C \rightarrow A\}$ és legyen r az alábbi reláció.

r	A	B	C
	a	c	e
	a	d	f
	b	c	g
	b	d	h

s	A	B
	a	c
	a	d
	b	c
	b	d

t	B	C
	c	e
	d	f
	c	g
	d	h

$s \bowtie t$	A	B	C
	a	c	e
	a	c	g
	a	d	f
	a	d	h
	b	c	e
	b	c	g
	b	d	f
	b	d	h

Ez a példa mutatja, hogy $r \neq s(A, B) \bowtie t(B, C)$, azaz ez a felbontás nem lesz jó. De $r = s'(A, C) \bowtie t'(B, C)$, majd látjuk.

r és $m_\rho(r)$ viszonya

Tétel

- (i) $r \subseteq m_\rho(r)$
- (ii) $r_i = \pi_{R_i}(m_\rho(r))$
- (iii) $m_\rho(m_\rho(r)) = m_\rho(r)$

Bizonyítás: Ha t egy sor, akkor $\pi_{R_i}(t)$ helyett $t[R_i]$ -t írunk.

- (i) Ha t egy sor r -ben, akkor t minden vetülete benne van a megfelelő $t[R_i]$ -ben, ezek össze is illenek, így $m_\rho(r)$ -ben is szerepelni fog t .
- (ii) $r \subseteq m_\rho(r) \implies r_i = \pi_{R_i}(r) \subseteq \pi_{R_i}(m_\rho(r))$.
Ha $t \in m_\rho(r)$, akkor ez természetes illesztéssel jött létre, r_i -beli sorokból, így levetítve R_i -re épp r_i egy sorát kapjuk.

$$m_\rho(m_\rho(r)) = m_\rho(r)$$

$$m_\rho(r) = \bowtie_{i=1}^k r_i = \bowtie_{i=1}^k \pi_{R_i}(r)$$

$$m_\rho(m_\rho(r)) = \bowtie_{i=1}^k \pi_{R_i}(m_\rho(r)) \stackrel{(ii)}{=} \bowtie_{i=1}^k r_i = m_\rho(r)$$

Megjegyzések

- (i) szerint a szétszedés és összerakás után vagy pont r -t kapom meg, vagy többet kapok, kevesebb sor nem lehet.
- Ha $r \neq m_\rho(r)$, akkor ez nem egy túl hasznos felbontás. De ennél több is igaz: ebben az esetben teljesen reménytelen a felbontásból visszaszerezni r -t: mivel (ii) szerint r és $m_\rho(r)$ (függőleges) vetületei ugyanazok, ezért ha $r \neq m_\rho(r)$, akkor *van két olyan reláció (r és $m_\rho(r)$), aminek a vetületei ugyanazok \implies a vetületekből nem lehet visszaállítani r -et (nem lehet eldönteni, hogy r vagy $m_\rho(r)$ volt).*
- Következmény: ha $r \neq m_\rho(r)$, akkor sehogyan se lehet visszahozni r -t a vetületekből.

Hûséges felbontás

Tehát az a kérdés, hogy mik azok a felbontásai egy (R, F) sémának, amik esetén tetszőleges (R, F) -re illeszkedő r relációra $r = m_\rho(r)$

Definíció

Adott (R, F) . Ennek ρ felbontása **hûséges (veszteségmentes, lossless)**, ha minden (R, F) -re illeszkedő r relációra $r = m_\rho(r)$.

Példa

Legyen (R, F) a következő: $R(A, B, C)$, $F = \{C \rightarrow A\}$ és legyen r az alábbi reláció.

r	A	B	C
	a	c	e
	a	d	f
	b	c	g
	b	d	h

s	A	B
	a	c
	a	d
	b	c
	b	d

t	B	C
	c	e
	d	f
	c	g
	d	h

$s \bowtie t$	A	B	C
	a	c	e
	a	c	g
	a	d	f
	a	d	h
	b	c	e
	b	c	g
	b	d	f
	b	d	h

Ez a példa mutatja, hogy $r \neq s(A, B) \bowtie t(B, C)$, azaz ez a felbontás nem hűséges. De $r = s'(A, C) \bowtie t'(B, C)$, majd látjuk.

Hûséges felbontás két részre

Hogyan biztosíthatja F , hogy a felbontás hûséges legyen?

Tétel

Az (R, F) séma $\rho = (R_1, R_2)$ felbontása hûséges \iff vagy

(a) $F \models R_1 \cap R_2 \rightarrow R_1 \setminus R_2$, vagy

(b) $F \models R_1 \cap R_2 \rightarrow R_2 \setminus R_1$.

Példa

$R(\text{TNÉV}, \text{TERMELŐ}, \text{ÁR}, \text{CÍM})$

$F = \{\text{TERMELŐ} \rightarrow \text{CÍM}; \text{TNÉV}, \text{TERMELŐ} \rightarrow \text{ÁR}\}$

$\rho = (\{\text{TERMELŐ}, \text{CÍM}\} ; \{\text{TNÉV}, \text{TERMELŐ}, \text{ÁR}\})$

$R_1 \cap R_2 = \{\text{TERMELŐ}\}, \quad R_1 \setminus R_2 = \{\text{CÍM}\}, \quad R_2 \setminus R_1 = \{\text{TNÉV}, \text{ÁR}\}$

$(\text{TERMELŐ})^+(F) = \{\text{TERMELŐ}, \text{CÍM}\} \supseteq R_1 \setminus R_2 \implies \text{hûséges}$

$\rho = (\{\text{TNÉV}, \text{TERMELŐ}\}; \{\text{TNÉV}, \text{CÍM}, \text{ÁR}\})$

$R_1 \cap R_2 = \{\text{TNÉV}\}, \quad R_1 \setminus R_2 = \{\text{TERMELŐ}\}, \quad R_2 \setminus R_1 = \{\text{CÍM}, \text{ÁR}\}$

$\implies (\text{TNÉV})^+(F)$ nem tartalmazza egyiket sem \implies nem hûséges

Bizonyítás egyik iránya

Bizonyítás: Tegyük fel, hogy $F \models R_1 \cap R_2 \rightarrow R_1 \setminus R_2$, belátjuk, hogy a felbontás hûséges. (Ha a másik igaz, ugyanígy.)

Legyen r egy tetszõleges reláció, $s = m_\rho(r)$. Elég belátni, hogy $s \subseteq r$, hiszen $r \subseteq s$ mindig igaz. Azaz, lássuk be, hogy ha t sora s -nek, akkor r -nek is.

Ha t sora s -nek, akkor $\exists u_1, u_2$ sorai r -nek, hogy $t[R_1] = u_1[R_1]$ és $t[R_2] = u_2[R_2]$.

$$\implies u_1[R_1 \cap R_2] = t[R_1 \cap R_2] = u_2[R_1 \cap R_2]$$

de ha két sor megegyezik a metszeten, akkor a feltétel miatt $R_1 \setminus R_2$ -n is \implies egyeznek az egész R_1 -en $\implies u_2$ és t egyeznek R_1 -en.

$\implies t = u_2$, hiszen R_1 -en a fenti miatt, R_2 -n a feltevés miatt egyeznek.

Bizonyítás másik iránya

Bizonyítás: Belátjuk, hogy ha $R_1 \setminus R_2 \not\subseteq (R_1 \cap R_2)^+(F)$ és $R_2 \setminus R_1 \not\subseteq (R_1 \cap R_2)^+(F)$, akkor ρ nem hûséges.

Legyen r a következõ kétsoros reláció:

r	R_1			R_2			$(R_1 \cap R_2)^+(F)$			$R_1 \cap R_2$					
t_1	0	0	0	1	1	1	1	1	1	1	1	1
t_2	1	1	1	1	1	1	1	1	1	0	0	0

A feltétel miatt a két szélsõ rész nem üres, ott nem egyezik meg a két sor. r -ben igazak F függõségei (lásd következõ slide).

Viszont $m_\rho(r) \not\supseteq r$, hiszen $m_\rho(r)$ -ben a csupa 1 sor is benne van.

Miért igazak F függései az előző slide r -jében?

Vegyünk egy $U \rightarrow V$ függést F -ben. Két lehetőség van:

- $U \subseteq (R_1 \cap R_2)^+$: ekkor V is benne van $(R_1 \cap R_2)^+$ -ben az $(R_1 \cap R_2)^+$ -t kiszámoló lezárási algoritmus miatt. Ekkor r két sora megegyezik U , de r definíciója miatt V -n is, vagyis a függés teljesül.
- $U \not\subseteq (R_1 \cap R_2)^+$: ekkor nincs két olyan sor r -ben, ami megegyezne U -n, vagyis a függés teljesül.

Hûségesség ellenõrzése általában

Adott (R, F) és $\rho = (R_1, \dots, R_k)$, ahol $R = A_1, \dots, A_n$.

Hogyan tudjuk eldõniteni, hogy hûséges-e a felbontás?

Készítünk egy $k \times n$ -es táblázatot:

	A_1	...	A_j	...	$A_{j'}$...	A_n
R_1							
\vdots							
R_i			a_j		$b_{ij'}$		
\vdots							
R_k							

- Kezdetben az (i, j) helyre a_j -t írunk, ha $A_j \in R_i$ és $b_{ij'}$ -t, ha $A_j \notin R_i$.

Táblázatos teszt

- Veszünk egy tetszőleges $X \rightarrow Y \in F$ függést.
Ha két sor megegyezik X -en, akkor egyenlővé tesszük Y -on is az alábbi módon:
 - Ha valahol a_j és b_{ij} van, akkor a b_{ij} -t a_j -ra cseréljük.
 - Ha b_{kj} és b_{lj} van, akkor az egyiket átírjuk a másikra.
- Ezt minden függésre megcsináljuk tetszőleges sorrendben, szükség esetén többször is.

Tétel

ρ pontosan akkor hûséges ha a végén lesz csupa a sor.

Nem bizonyítjuk.

Példa a táblázatos tesztre

$R(ABCD) \quad F = \{A \rightarrow C; C \rightarrow B; B \rightarrow D\} \quad \rho = (AB, BC, ACD)$

	A	B	C	D
R_1	a_1	a_2	$b_{13} \rightarrow a_3$ ⁱ	b_{14}
R_2	b_{21}	a_2	a_3	b_{24}
R_3	a_1	$b_{32} \rightarrow a_2$ ⁱⁱ	a_3	a_4

ⁱ $A \rightarrow C$ miatt

ⁱⁱ $C \rightarrow B$ miatt

Lett csupa a sor \implies hûséges felbontás

Hûséges felbontás

Tétel

Adott (R, F) , $\rho = (R_1, \dots, R_k)$ az R hûséges felbontása és $\sigma = (S_1, \dots, S_m)$ az R_1 hûséges felbontása (azaz R_1 -et tovább bontjuk). Ekkor $\tau = (S_1, \dots, S_m, R_2, \dots, R_k)$ hûséges felbontása R -nek.

Nem bizonyítjuk.