Adatbáziskezelés
Normálformák

Csima Judit

BME, VIK,
Számítástudományi és Információelméleti Tanszék

2017. november 8.

Felbontások

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

Definíció

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

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

(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?

r és $m_\rho(r)$ viszonya

Tétel

(i) $r \subseteq m_\rho(r)$

(ii) $r_i = \pi_{R_i}(m_\rho(r))$

(iii) $m_\rho\left(m_\rho(r)\right) = m_\rho(r)$

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 megfelelõ $t[R_i]$-ben, ezek össze is illenek, így $m_\rho(r)$-ben is szerepelni fog t.

(ii) $r \subseteq m_\rho(r) \implies r_i = \pi_{R_i}(r) \subseteq \pi_{R_i}\left(m_\rho(r)\right)$.
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.

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

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

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 $m_\rho(r)$ (függõleges) vetületei ugyanazok, ezért ha $r \neq m_\rho(r)$, akkor *van két olyan reláció (r és $m_\rho(r)$), aminek a vetületei ugyanazok $\Longrightarrow$ a vetületekbõl 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 vetületekbõl.

Hûséges felbontás

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

Definíció

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

Példa

Legyen (R, F) a következõ: $R(A, B, C)$, $F = \{C \to A\}$ 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 hûséges. 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 $\Longleftrightarrow$ vagy

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

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

Példa

R(TNÉV, TERMELÕ, ÁR, CÍM)
$F = \{$TERMELÕ $\to$ CÍM; TNÉV, TERMELÕ $\to$ ÁR$\}$

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

$R_1 \cap R_2 = \{$TERMELÕ$\}$, $R_1 \setminus R_2 = \{$CÍM$\}$, $R_2 \setminus R_1 = \{$TNÉV, ÁR$\}$

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

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

$R_1 \cap R_2 = \{$TNÉV$\}$, $R_1 \setminus R_2 = \{$TERMELÕ$\}$, $R_2 \setminus R_1 = \{$CÍM, ÁR$\}$

$\implies ($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 $t[R_2] = u_2[R_2]$.
$\Longrightarrow u_1[R_1 \cap R_2] = t[R_1 \cap R_2] = u_2[R_1 \cap R_2]$
de ha két sor megegyezik a metszeten, akkor a feltétel miatt $R_1 \setminus R_2$-n is $\Longrightarrow$ egyeznek az egész R_1-en $\Longrightarrow u_2$ és t egyeznek R_1-en.
$\Longrightarrow t = u_2$, 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_1 R_2 $(R_1 \cap R_2)^+(F)$ $R_1 \cap R_2$

r													
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 (mint a teljességi tételnél).
Viszont $m_\rho(r) \supsetneq r$, hiszen $m_\rho(r)$-ben a csupa 1 sor is benne van.

Hûségesség ellenõrzése általában

Adott (R, F) és $\rho = (R_1, \ldots, R_k)$, ahol $R = A_1, \ldots, A_n$.
Hogyan tudjuk eldönteni, hogy hûséges-e a felbontás?
Készítünk egy $k \times n$-es táblázatot:

	A_1	$\ldots$	A_j	$\ldots$	$A_{j'}$	$\ldots$	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 tetszõleges $X \to Y \in F$ függést.
  Ha két sor megegyezik X-en, akkor egyenlõvé 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 tetszõleges sorrendben, szükség esetén többször is.

Tétel

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

Nem bizonyítjuk.

Példa a táblázatos tesztre

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

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

[i] $A \to C$ miatt
[ii] $C \to B$ miatt

Lett csupa a sor $\Longrightarrow$ hûséges felbontás

Hûséges felbontás

Tétel

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

Nem bizonyítjuk.

Tétel

Ha ρ hûséges és $\sigma \supseteq \rho$ (σ-ban több komponens van), akkor σ is hûséges.

Bizonyítás: $r \subseteq m_\sigma(r) \subseteq m_\rho(r) = r$

A középsõ tartalmazás azért igaz, mert a keresztszorzatból szigorúbb feltételek szerint válogatunk.

$\implies r = m_\sigma(r)$

Normálformák

Definíció

Egy $X \to Y$ függés triviális, ha $Y \subseteq X$. (Mert ezek a függések nem hordoznak sok infót, mindig igazak.)

Definíció (Boyce–Codd normálforma)

Az (R, F) relációs séma BCNF-ben van, ha tetszõleges nemtriviális $X \to A \in F^+$ függés esetén X szuperkulcs.

Azaz csak olyan függések vannak, hogy a szuperkulcs mindent meghatároz.

Miért jó a BCNF séma?

Ha $C \rightarrow B$; $B \rightarrow A$ teljesülne, de $B \rightarrow C$ nem, akkor ugyanaz a B érték több C érték mellett is elõfordulhatna, de minden példánynál ugyanazt az A értéket is tároljuk $\Longrightarrow$ **redundancia**.

2 attribútumos relációk vs. BCNF

Állítás

≤ 2 *attribútumos reláció mindig BCNF.*

Bizonyítás: Ha $A \to B \implies A$ kulcs. Ha $B \to A \implies B$ kulcs, azaz minden függés baloldala szuperkulcs. Nincs tehát olyan nemtriviális függés, ami meg bírná sérteni a BCNF követelményét.

BCNF-ség ellenőrzése

Mivel F^+-ra követeljük meg a feltételt, nehéz ellenõrizni.
DE:

Tétel

Ha (R, F) nem BCNF, akkor van olyan $X \rightarrow Y \in \mathbf{F}$, amely jobboldalának valamely A attribútumára $X \rightarrow A$ nemtriviális és X nem szuperkulcs. (Az ilyen $X \rightarrow A \in F^+$.)

Azaz elég F-et vizsgálnom!!

Bizonyítás

Bizonyítás: Ha (R, F) nem BCNF, akkor van $U \to B \in F^+$, hogy U nem szuperkulcs és $B \notin U$. $\implies B \in U^+(F) \implies U \subsetneq U^+(F)$
Az algoritmus, ami $U^+(F)$-et számolja, el tud indulni $\implies \exists V \to W \in F$, melyre $V \subseteq U$, $W \not\subseteq U \implies V \to W$ jó lesz $X \to Y$-nak.
Ugyanis V nem szuperkulcs, hiszen $V \subseteq U$ és U nem szuperkulcs.
$W \not\subseteq U \implies \exists A \in W \setminus U \subseteq W \setminus V$, így $V \to W$ nem triviális.

Ez jelentõsen könnyíti az ellenõrzést, csak F függõségeit kell végignézni, nem F^+-ét.