L-relacijska algebra

L-relational algebra

Melita Hajdinjak

Univerza v Ljubljani, Fakulteta za elektrotehniko / Tržaška cesta 25, 1000 Ljubljana

E-Mail: melita.hajdinjak@fe.uni-lj.si

* Avtor za korespondenco; Tel.: +386-1-4768-385; Fax: + 386-1-4768-316

Povzetek: V članku se osredotočimo na pozitivni K-relacijski model za označene relacije, v katerem U-tericam prirejamo elemente izbranega komutativnega polkolobarja K. Predlagamo uvedbo nove operacije v K, imenovane negacija, ki omogoča definicijo razlike K-relacij, tj. edine osnovne relacijske operacije, ki v pozitivni K-relacijski algebri manjka. Obstoj negacije zagotavljajo posebne mreže, imenovane De Morganovi okvirji. Zahteva, da naj bo K De Morganov okvir, ima še en pozitiven učinek −v dobljeni L-relacijski algebri veljajo vse (pozitivne) relacijske identitete, ki jih poznamo iz klasične relacijske algebre, tudi tiste, ki v K-relacijski algebri ne veljajo.

Ključne besede: K-relacija; L-relacija; negacija; De Morganov okvir; relacijska algebra

Abstract: In this article we focus on the positive K-relation model for annotated relations in which each U-tuple is mapped to an element of a commutative semiring, K. We propose a new operation in K, called negation, which induces a difference operation on K-relations, i.e., the only basic relational operation not interpreted in the positive K-relational algebra. Because we cannot define negation in all semirings, we move from the commutative semirings to the De Morgan frames, a lattice structure with negation. This has another positive consequence – the obtained L-relational algebra satisfies all the classical relational identities, including those that are not satisfied by the positive K-relational algebra.

Keywords: K-relation; L-relation; negation; De Morgan frame; relational algebra

1. Uvod

Predpostavimo, da imamo na razpolago neke osnovne domene oziroma množice elementov, ki jih bomo označevali s τ. Primera sta množica naravnih števil in množica nizov. Naj bo *shema*

$$U = \{a_1 : \tau_1, \ldots, a_n : \tau_n\}$$

preslikava, ki atributu a_i priredi domeno τ_i, torej $U(a_i) = \tau_i$ za $i = 1, \ldots, n$. Naj bo *U-terica*

$$t = \{a_1 : v_1, \ldots, a_n : v_n\}$$

preslikava, ki atributu a_i priredi element $v_i \in \tau_i$, torej $t(a_i) = v_i$ za $i = 1, \ldots, n$. Množico vseh U-teric imenujmo $U - \mathrm{Tup}$. *Označena relacija nad U* bomo imenovali preslikavo oblike $A : U - \mathrm{Tup} \to K$, ki vsaki U-terici priredi nek element izbrane domene K. Klasična relacija, ki jo običajno pojmujemo kot množico U-teric in jo predstavimo s tabelo, je poseben primer označene relacije. Dobimo jo, če izberemo K = {0,1} in definiramo $A(t) = 1$, če $t \in A$, ter $A(t) = 0$ sicer.

Naj bo sedaj K poljuben *polkolobar*, torej algebrajska struktura $K = (K, +, *, 0, 1)$ z dvema dvomestnima operacijama (+ in *) ter različnima elementoma 0 in 1, tako da velja naslednje:

- $(K, +, 0)$ je komutativen monoid z enoto 0,
- $(K, *, 1)$ je monoid z enoto 1,
- produkt (*) je distributiven glede na vsoto (+) in
- za vsak $a \in K$ velja $0 * a = a * 0 = 0$.

Polkolobar K imenujemo *komutativen*, če je komutativen monoid $(K,*,1)$.

V članku se bomo osredotočili na tako imenovane *K-relacije* in pozitivno *K-relacijsko algebro* [1], katerih namen je modelirati različne oblike *označenih relacij*.

2. Relacijska algebra na K-relacijah

Green, Karvounarakis in Tannen [1] so predlagali naslednjo razširitev klasičnih relacij.

Definicija 1. [K-relacija] Naj bo $K=(K,+,*,0,1)$ komutativen polkolobar. *K-relacija nad U*, kjer je $U=\{a_1:\tau_1,\dots,a_n:\tau_n\}$ relacijska shema, je označena relacija $A:U-\mathrm{Tup}\to K$, ki ima končen nosilec, tj.

$$\mathrm{supp}(A)=\{t:A(t)\neq 0\}$$

je končna množica.

Definicija 2. [Pozitivna K-relacijska algebra] Naj bo $K=(K,+,*,0,1)$ komutativen polkolobar. Operacije *pozitivne K-relacijske algebre* so naslednje:

- *Prazna relacija*: Za vsako shemo U imamo prazno K-relacijo Ø: $U-\mathrm{Tup}\to K$, za katero za vsak t velja

$$Ø(t) = 0.$$

- *Unija*: Če sta $A,B:U-\mathrm{Tup}\to K$, potem je unija $A\cup B:U-\mathrm{Tup}\to K$ definirana kot

$$(A\cup B)(t)=A(t)+B(t).$$

- *Projekcija:* Če je $A:U-\mathrm{Tup}\to K$ in je V podmnožica množice U, je projekcija $\pi_V A:V-\mathrm{Tup}\to K$ definirana kot

$$(\pi_V A)(t)=\sum_{t'\downarrow V=t \text{ in } A(t')\neq 0} A(t'),$$

kjer je $t'\downarrow V$ skrčitev U-terice t' na domeno V.

- *Izbira*: Če je $A:U-\mathrm{Tup}\to K$ in predikatna funkcija **P** slika U-terico v 0 ali 1, potem je izbira $\sigma_{\mathbf{P}}A:U-\mathrm{Tup}\to K$ definirana kot

$$(\sigma_{\mathrm{P}}A)(t)=A(t)*\mathbf{P}(t).$$

- *Stik*: Če sta $A:U_1-\mathrm{Tup}\to K$ in $B:U_2-\mathrm{Tup}\to K$, potem je njun stik $A\otimes B:U_1\cup U_2\to K$ definiran kot

$$(A\otimes B)(t)=A(t\downarrow U_1)*B(t\downarrow U_2).$$

- *Preimenovanje*: Če je $A:U-\mathrm{Tup}\to K$ in je $\beta:U\to U'$ bijektivna preslikava, potem definiramo preimenovano K-relacijo $\rho_\beta A:U'-\mathrm{Tup}\to K$ kot

$$(\rho_\beta A)(t)=A(t\circ\beta).$$

K-relacijska algebra posplošuje številne predloge modelov označenih relacij. Če za K izberemo Boolov (pol)kolobar, dobimo klasično relacijsko algebro, če izberemo polkolobar celih števil, dobimo algebro na vrečah [2], če izberemo ustrezni polkolobar na končni množici dogodkov, dobimo Fuhr-Rölleke-Zimányi verjetnostno algebro [3], če izberemo polkolobar c-tabel, dobimo Imielinksi-Lipski algebro [4], in če izberemo provenienčni polkolobar, dobimo provenienčno algebro [1].

V Definiciji 2 manjka ena osnovna relacijska operacija – razlika. Green, Karvounarakis in Tannen [1] so ta (negativni) del pustili odprt, saj v komutativnih polkolobarjih niso znali definirati ustrezne operacije, ki bi posplošila razliko, kot jo poznamo v klasični relacijski algebri [5]. Pozneje sta Geerts in Poggi [6] predlagala, da bi razliko definirali kot *monus*, zaradi česar bi bilo treba (komutativne) polkolobarje skrčiti na (komutativne) polkolobarje z monusom, imenovane *m-polkolobarji*. A ta definicija razlike označenih relacij se je izkazala za neustrezno v primeru mehkih polkolobarjev in mehkih relacijskih algeber [7].

Izrek 1. [Identitete na K-relacijah] Pozitivna K-relacijska algebra ima naslednje lastnosti:

- unija je asociativna in komutativna operacija z enoto Ø,
- izbira je distributivna glede na unijo in stik,
- stik je asociativna in komutativna operacija, ki je distributivna glede na unijo,
- projekcija je distributivna glede na unijo in stik,
- izbira in projekcija komutirata,
- če je **P**(t) = 1 za vse t, je $\sigma_{\mathrm{P}}A=A$, in če je **P**(t) = 0 za vse t, je $\sigma_{\mathrm{P}}A=Ø$,
- stik s prazno relacijo je prazna relacija,
- projekcija prazne relacije je prazna relacija.

V pozitivni K-relacijski algebri pa ne velja $A\cup A=A$ in $A\otimes A=A$. Unija in samostik sta sicer idempotenta v klasičnem modelu [5], ne pa tudi v provenienčni algebri [1] in algebri na vrečah [2].

3. Relacijska algebra na L-relacijah

Najprej predlagamo manjšo posplošitev predikatne funkcije iz pozitivnega K-relacijskega modela, potem pa se osredotočimo na problem definicije razlike označenih relacij. S komutativnih polkolobarjev K preidemo na *De Morganove okvirje* L, ki vsebujejo *negacijo*, s pomočjo katere definiramo razliko dobljenih *L-relacij*. Na ta način dosežemo tudi veljavnost vseh (pozitivnih) relacijskih identitet, ki jih poznamo iz klasičnega relacijskega modela.

3.1. Predikatna funkcija

Namesto da bi se omejili le na 0 (neizpolnjevanje pogoja) in 1 (popolno izpolnjevanje pogoja), kot so to naredili Green, Karvounarakis in Tannen [1], bomo dovolili tudi druge vrednosti predikatne funkcije **P**, ki bodo odražale različne stopnje pripadnosti oziroma izpolnjevanja pogoja.

Definicija 3. [Posplošena izbira] Predikatna funkcija $\mathbf{P}: U-\mathrm{Tup} \rightarrow K$ naj U-terico slika v poljubni element komutativnega polkolobarja K.

- *Izbira*: Če je $A: U-\mathrm{Tup} \rightarrow K$ in $\mathbf{P}: U-\mathrm{Tup} \rightarrow K$, izbiro $\sigma_{\mathbf{P}} A: U-\mathrm{Tup} \rightarrow K$ definiramo kot

$$(\sigma_{\mathrm{P}} A)(t)=A(t) * \mathbf{P}(t).$$

V nadaljevanju, ko bomo govorili o pozitivni K-relacijski algebri, bomo mislili to posplošitev.

Naš cilj je v splošen model označenih relacij dodati tudi mehke relacije [7]. V mehkih relacijah so U-terice največkrat označene z Boolovima vrednostma (0 in 1), z vrednostmi z enotskega intervala [0,1] ali z vrednostmi s kompaktificiranega intervala [0,∞]. Vsaka izmed teh množic značk tvori komutativen polkolobar:

- *Boolov polkolobar* je $K=(\{0,1\}, \vee, \wedge, 0,1)$,
- *mehki polkolobar* je $K=([0,1], \max, \min, 0,1)$,
- *polkolobar razdalj je* $K=([0, \infty], \min, \max, \infty, 0)$.

Z elementi komutativnega polkolobarja lahko izražamo podobnost U-teric. Tako z največjim elementom 1 izrazimo največjo podobnost (enakost), z najmanjšim elementom 0 pa najmanjšo podobnost (različnost). V polkolobarju razdalj smo interval [0, ∞] obrnili, tako da 0 pomeni največjo podobnost oziroma najmanjšo razdaljo, ∞ pa najmanjšo podobnost oziroma največjo (nepremostljivo) razdaljo.

3.2. Razlika

Razliko označenih relacij bomo definirali s pomočjo negacije.

Definicija 5. [Negacija] Operacijo $\neg: L \rightarrow L$ na delno urejeni množici *L*, ki obrača urejenost ($x \leq y \Rightarrow \neg y \leq \neg x$) in je involutivna ($\neg\neg x=x$), imenujemo *negacija*. Komutativni polkolobar, ki je delno urejen in ima negacijo, imenujemo *n-polkolobar*.

Definicija 6. [Razlika] Naj bo *L* = (*L*,+,*,0,1,¬) n-polkolobar.

- *Razlika*: Če sta $A, B: U-\mathrm{Tup} \rightarrow L$, potem je razlika $A-B: U-\mathrm{Tup} \rightarrow L$ definirana kot

$$(A-B)(t)=A(t) * \neg B(t).$$

Poglejmo, kakšno definicijo razlike dobimo v nekaterih primerih:

- V Boolovem n-polkolobarju dobimo $(A-B)(t)=0$, če je $B(t)=1$, in $(A-B)(t)=A(t)$, če je $B(t)=0$.
- V mehkem n-polkolobarju dobimo $(A-B)(t)=\min\{A(t), 1-B(t)\}$, kar se ujema s klasično definicijo razlike mehkih relacij [7].
- V n-polkolobarju razdalj dobimo $(A-B)(t)=A(t)$, če je $B(t)=\infty$, $(A-B)(t)=\max\{A(t), 1/B(t)\}$, če je $B(t) \neq \infty$ in $B(t) \neq 0$, in $(A-B)(t)=\infty$, če je $B(t)=0$.
- V n-polkolobarju na končni množici dogodkov dobimo $(A-B)(t)=A(t) \cap(\Omega-B(t))$, kar se ujema z definicijo razlike v Fuhr-Rölleke-Zimányi verjetnostni algebri [3].
- V n-polkolobarju c-tabel dobimo $(A-B)(t)=A(t) \wedge \neg B(t)$, kar se ujema z razliko v Imielinksi-Lipski algebri [4].

Z uvedbo negacije in prehodom na n-polkolobarje smo sicer zagotovili obstoj razlike na mehkih relacijah, na verjetnostnih relacijah in v Imielinksi-Lipski algebri, ne pa tudi v provenienčni algebri in algebri na vrečah.

3.3. L-relacije

Opazimo, da so Boolov polkolobar, mehki polkolobar, polkolobar razdalj, verjetnostni polkolobar in polkolobar na c-tabelah vsi polne mreže z operacijo supremum definirano kot + in operacijo infimum definirano kot *. Obe operaciji na mrežah sta idempotentni. Znan primer polnih mrež z negacijo so *De Morganovi okvirji.* Tako imenujemo polno mrežo $L=(L,\vee,\wedge,0,1,\neg)$ z negacijo, v kateri so končni infimumi distributivni glede na supremume. Vse prej omenjene polne mreže so tudi De Morganovi okvirji.

Definicija 7. [L-relacija] Naj bo $L=(L,\vee,\wedge,0,1,\neg)$ De Morganov okvir. L-*relacija nad U*, kjer je $U=\{a_1:\tau_1,\dots,a_n:\tau_n\}$ relacijska shema, je označena relacija $A:U-\mathrm{Tup}\rightarrow L$.

Ker v polnih mrežah obstaja supremum poljubne množice, obstaja tudi projekcija L-relacij. Zaradi tega sedaj ne zahtevamo, da je nosilec L-relacij končen.

Definicija 8. [L-relacijska algebra] Naj bo $L=(L,\vee,\wedge,0,1,\neg)$ De Morganov okvir. Operacije L-*relacijske algebre* so naslednje:

- *Prazna relacija*: Za vsako shemo U imamo prazno L-relacijo Ø: $U-\mathrm{Tup}\rightarrow L$, za katero za vsak t velja

  $$Ø(t) = 0.$$

- *Unija*: Če sta $A,B:U-\mathrm{Tup}\rightarrow L$, potem je unija $A\cup B:U-\mathrm{Tup}\rightarrow L$ definirana kot

  $$(A\cup B)(t)=A(t)\vee B(t).$$

- *Projekcija:* Če je $A:U-\mathrm{Tup}\rightarrow L$ in je V podmnožica množice U, je projekcija $\pi_V A:V-\mathrm{Tup}\rightarrow L$ definirana kot

  $$(\pi_V A)(t)=\bigvee_{t'\downarrow V=t \text{ in } A(t')\neq 0} A(t'),$$

  kjer je $t'\downarrow V$ skrčitev U-terice t' na domeno V.

- *Izbira*: Če je $A:U-\mathrm{Tup}\rightarrow L\rightarrow$ in predikatna funkcija $\mathbf{P}:U-\mathrm{Tup}\rightarrow L$, potem je izbira $\sigma_{\mathbf{P}}A:U-\mathrm{Tup}\rightarrow L$ definirana kot

  $$(\sigma_{\mathrm{P}}A)(t)=A(t)\wedge\mathbf{P}(t).$$

- *Stik*: Če sta $A:U_1-\mathrm{Tup}\rightarrow L$ in $B:U_2-\mathrm{Tup}\rightarrow L$, potem je njun stik $A\otimes B:U_1\cup U_2\rightarrow L$ definiran kot

  $$(A\otimes B)(t)=A(t\downarrow U_1)\wedge B(t\downarrow U_2).$$

- *Preimenovanje*: Če je $A:U-\mathrm{Tup}\rightarrow L$ in je $\beta:U\rightarrow U'$ bijektivna preslikava, potem definiramo preimenovano L-relacijo $\rho_\beta A:U'-\mathrm{Tup}\rightarrow L$ kot

  $$(\rho_\beta A)(t)=A(t\circ\beta).$$

- *Razlika*: Če sta $A,B:U-\mathrm{Tup}\rightarrow L$, potem je razlika $A-B:U-\mathrm{Tup}\rightarrow L$ definirana kot

  $$(A-B)(t)=A(t)\wedge\neg B(t).$$

Ker sta infimum in supremum idempotentni operaciji v mrežah, sta idempotentna tudi unija in (samo)stik L-relacij. Spomnimo se, da to ni veljalo za K-relacije. Posledično v L-relacijski algebri veljajo vse (pozitivne) relacijske identitete.

Izrek 2. [Identitete na L-relacijah] L-relacijska algebra ima naslednje lastnosti:

- unija je asociativna, komutativna in idempotentna operacija z enoto Ø,
- izbira je distributivna glede na unijo in stik,
- stik je asociativna, komutativna in idempotentna operacija, ki je distributivna glede na unijo,
- projekcija je distributivna glede na unijo in stik,
- izbira in projekcija komutirata,
- če je $\mathbf{P}(t) = 1$ za vse t, je $\sigma_{\mathrm{P}}A=A$, in če je $\mathbf{P}(t) = 0$ za vse t, je $\sigma_{\mathrm{P}}A=$ Ø,
- stik s prazno relacijo je prazna relacija,
- projekcija prazne relacije je prazna relacija.

Identitete na L-relacijah, ki vključujejo tudi razliko, so podane in dokazane v [9].

4. Zaključki

Z uvedbo negacije v komutativne polkolobarje smo znali definirati tudi razliko označenih relacij. Pokazali smo, da se tako definirana razlika ujema z razliko, kot jo poznamo v klasični relacijski algebri, v mehkih relacijskih algebrah, v verjetnostni relacijski algebri in v relacijski algebri na c-tabelah. Pri tem smo komutativne polkolobarje K zamenjali z De Morganovi okvirji L in prešli s K-relacij na L-relacije. V dobljeni L-relacijski algebri veljajo vse (pozitivne) relacijske identitete, ki jih poznamo iz klasične relacijske algebre, tudi tiste, ki v K-relacijski algebri niso veljale.

Na žalost pa naša L-relacijska algebra ne posplošuje niti prevenienčne algebre niti relacijske algebre na vrečah; ta dva primera ne vsebujeta negacije. V nadaljnjem delu se zato nameravamo posvetiti temu problemu. Poiskali bi radi splošnejši pojem razlike označenih relacij ali pa poskušali provenienco obravnavati na drugačen način.

Literatura

1. Green, T.J.; Karvounarakis, G.; Tannen, V. Provenance semirings. *Proceedings of the 26th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, Peking, Kitajska.* **2007**, 31-40.
2. Montagna, F.; Sebastiani, V. Equational Fragments of Systems for Arithmetic. Algebra Universalis **2001**, 46 (3), 417-441.
3. Suciu, D. Probabilistic databases. SIGACT News **2008**, 39 (2), 111-124.
4. Imielinski, T.; Lipski, W. Incomplete information in relational databases. Journal of the ACM **1984**, 31 (4), 761-791.
5. Ullman, J.D. Principles of Database and Knowledge-Base Systems, Volume I; Publisher: Computer Science Press, Inc., 1988.
6. Geerts, F.; Poggi, A. On Database Query Languages for K-relations. Journal of Applied Logic **2010**, 8, 173-185.
7. Buckles, B.P.; Petry, F. A fuzzy Representation of Data for Relational Databases. Fuzzy Sets and Systems **1982**, 7, 213-226.
8. Hajdinjak, M.; Bauer, A. Similarity Measures for Relational Databases. Informatica **2009**, 33 (2), 135-141.
9. Hajdinjak, M.; Bierman, G. Extending the Relational Algebra with Similarities. Poslano v Mathematical Structures in Computer Science **2011.**