

Formálna konceptová analýza

Ondrej Krídlo

Pavol Jozef Šafárik University in Košice

Prírodovedecká fakulta, Ústav informatiky



Formálna konceptová analýza

- Aplikovaná teória zväzov
- Rudolf Wille "Restructuring Lattice Theory" (1982, Darmstadt)
- *Reštrukturalizácia teórie zväzov je pokus o oživenie interpretácie tejto teórie čo najkonkrétnejšie, a týmto spôsobom podporiť lepšiu komunikáciu medzi teoretikmi a potenciálnymi používateľmi teórie zväzov.*
- korene je možné vystopovať v práci Charles S. Peirce-a a v texte Port-Royal Logic.

Formálny kontext

- objektovo atribútová tabuľka $\langle B, A, R \rangle$
- B je množina objektov
- A je množina atribútov
- $R \subseteq B \times A$ je relácia (vzťah) medzi objektmi a atribútmi

	M	F	I	Ch	B
Anna	•	•	•		
Boris	•		•		•
Cyril				•	•
Dana		•		•	•

Derivačné operátory

- $\uparrow: P(B) \rightarrow P(A)$ a $\downarrow: P(A) \rightarrow P(B)$
- $X \subseteq B \dots \uparrow(X) = \{a \in A \mid (\forall b \in X) (b, a) \in R\}$
- $\uparrow(X)$ je množina všetkých atribútov zdieľané všetkými objektami z množiny X
- $Y \subseteq A \dots \downarrow(Y) = \{b \in B \mid (\forall a \in Y) (b, a) \in R\}$
- $\downarrow(Y)$ je množina všetkých objektov, ktorí zdieľajú všetky atribúty z množiny Y

Formálny koncept

- dvojica (X, Y) taká, že $\uparrow(X) = Y$ a $\downarrow(Y) = X$
- inými slovami $X \times Y \subseteq R$ maximálny obdĺžnik pod R (po permutácii riadkov a stĺpcov)

	M	F	I	Ch	B
Anna	•	•	•		
Boris	•		•		•
Cyril				•	•
Dana		•		•	•

Formálny koncept

- dvojica (X, Y) taká, že $\uparrow(X) = Y$ a $\downarrow(Y) = X$
- inými slovami $X \times Y \subseteq R$ maximálny obdĺžnik pod R (po permutácii riadkov a stĺpcov)

	M	F	I	Ch	B
Anna	•	•	•		
Boris	•		•		•
Cyril				•	•
Dana		•		•	•

Formálny koncept

- dvojica (X, Y) taká, že $\uparrow(X) = Y$ a $\downarrow(Y) = X$
- inými slovami $X \times Y \subseteq R$ maximálny obdĺžnik pod R (po permutácii riadkov a stĺpcov)

	M	F	I	Ch	B
Anna	•	•	•		
Boris	•		•		•
Cyril				•	•
Dana		•		•	•

Formálny koncept

- dvojica (X, Y) taká, že $\uparrow(X) = Y$ a $\downarrow(Y) = X$
- inými slovami $X \times Y \subseteq R$ maximálny obdĺžnik pod R (po permutácii riadkov a stĺpcov)

	M	F	I	Ch	B
Anna	•	•	•		
Boris	•		•		•
Cyril				•	•
Dana		•		•	•

Formálny koncept

- dvojica (X, Y) taká, že $\uparrow(X) = Y$ a $\downarrow(Y) = X$
- inými slovami $X \times Y \subseteq R$ maximálny obdĺžnik pod R (po permutácii riadkov a stĺpcov)

	M	F	I	Ch	B
Anna	•	•	•		
Boris	•		•		•
Cyril				•	•
Dana		•		•	•

Formálny koncept

- dvojica (X, Y) taká, že $\uparrow(X) = Y$ a $\downarrow(Y) = X$
- inými slovami $X \times Y \subseteq R$ maximálny obdĺžnik pod R (po permutácii riadkov a stĺpcov)

	M	F	I	Ch	B
Anna	•	•	•		
Boris	•		•		•
Cyril				•	•
Dana		•		•	•

Matematické pozadie

- $(\uparrow, \downarrow)$ je Galoisova konexia medzi $(P(B), \subseteq)$ a $(P(A), \subseteq)$
- ak $X_1 \subseteq X_2 \subseteq B$ tak $\uparrow(X_1) \supseteq \uparrow(X_2)$
- ak $Y_1 \subseteq Y_2 \subseteq A$ tak $\downarrow(Y_1) \supseteq \downarrow(Y_2)$
- $X \subseteq \downarrow \uparrow(X)$ a $Y \subseteq \uparrow \downarrow(Y)$
- ekvivalentn pre vš. $X \subseteq B$ a $Y \subseteq A$ platí

$$X \subseteq \downarrow(Y) \iff Y \subseteq \uparrow(X)$$

- inými slovami $X \times Y \subseteq R$



Matematické pozadie

- spojenia $\downarrow\uparrow$ a $\uparrow\downarrow$ sú uzáverové operátory na $P(B)$ a $P(A)$
- $\downarrow\uparrow: P(B) \rightarrow P(B)$ je uzáverový operátor ak
- $X_1 \subseteq X_2 \subseteq B$ tak $\downarrow\uparrow(X_1) \supseteq \downarrow\uparrow(X_2)$
- $X \subseteq B$ platí $X \subseteq \downarrow\uparrow(X)$
- $X \subseteq B$ platí $\downarrow\uparrow\downarrow\uparrow(X) = \downarrow\uparrow(X)$

$X \subseteq \downarrow\uparrow(X)$ tak $\uparrow(X) \supseteq \uparrow(\downarrow\uparrow(X)) = \uparrow\downarrow(\uparrow(X)) \supseteq \uparrow(X)$

- formálny koncept (X, Y) , že $\uparrow(X) = Y$ a $\downarrow(Y) = X$
- dvojica uzavretých množín objektov a atribútov X a Y , že $X \times Y \subseteq R$
- „pomenovaniahodné“, koncept $(X, Y, \textit{meno})$

Crimes of London

- Barking and Dagenham, Barnet, Bexley, Brent, Bromley, Camden, Croydon, Ealing, Enfield, Greenwich, Hackney, Hammersmith and Fulham, Haringey, Harrow, Havering, Hillingdon, Hounslow, Islington, Kensington and Chelsea, Kingston upon Thames, Lambeth, Lewisham, Merton, Newham, Redbridge, Richmond upon Thames, Southwark, Sutton, Tower Hamlets, Waltham Forest, Wandsworth, Westminster
- Violence Against the Person, Theft and Handling, Drugs, Other Notifiable Offences, Robbery, Criminal Damage, Burglary, Sexual Offences, Fraud or Forgery
- početnosť výskytov v daných okresoch v rokoch 2008 - 2016

Crimes of London

Bromley	46759	69742	2637	4868	24039	27135	31	196
Croydon	67791	91437	4340	12645	31218	33376	55	270
Enfield	45036	70371	3293	9059	22487	30213	38	132
Hackney	56584	91118	3332	8975	17327	21450	46	143
Lambeth	72726	114899	4520	18408	26136	30199	70	137
Haringey	50943	83979	2971	10084	22272	28213	40	207
Havering	33968	52609	2358	3089	17252	21302	19	179
Hounslow	51404	70180	3963	4847	21407	21026	40	183
Lewisham	63652	70382	3809	10455	24810	24871	71	262
Greenwich	52897	64923	3598	5430	22755	20966	56	107
Islington	52975	107661	3675	8736	18354	22207	40	85
Redbridge	41430	71496	2619	7688	17543	26735	31	284
Southwark	68356	109432	4696	16153	24450	27980	40	321
Hillingdon	55264	80028	6504	5663	24485	26056	44	223
Wandsworth	45865	92523	3091	8398	19630	25533	47	161
Westminster	71448	277617	6148	15752	20405	29295	59	273
Tower_Hamlets	59993	87620	4268	10050	21593	21510	47	124



Úprava dát

- zločinnosť v každom okrese sme vydelili počtom obyvateľov
- následne sme pre každý zločin (stĺpec) našli jeho maximum a predelili ním všetky hodnoty v danom stĺpci
- tým sme dostali (normované) hodnoty iba z intervalu $[0, 1]$
- $r : B \times A \rightarrow [0, 1]$ fuzzy relácia medzi objektami a atribútami
- r je fuzzy podmnožina $B \times A$
- fuzzy (pod)množina umožňuje okrem striktných pravdivostných hodnôt, stupňov príslušnosti prvkov {true, false} alebo {1,0} aj vágne pojmy {možno, trochu, veľmi, menej, ...} alebo {0.75, 0.26578, ...}



Crimes of London

Barking and Dagenham	0.244355	0.671621	0.415252		
Barnet	0.226779	0.443924	0.218917	0.299566	
Bexley	0.161454	0.420482	0.24622	0.278407	0.116047
Brent	0.247407	0.640915	0.59568	0.378467	0.38048
Bromley	0.208749	0.464952	0.20776	0.295671	0.15341
Camden	0.514898	0.707992	0.565766	0.461333	
Croydon	0.217206	0.615884	0.333573	0.4353	0.404894
Ealing	0.26322	0.62573	0.368323	0.413992	0.264918
Enfield	0.21821	0.510975	0.29924	0.344106	0.330944
Greenwich	0.257611	0.701754	0.289105	0.431634	
Hackney	0.389837	0.737984	0.461457	0.422839	
Hammersmith and Fulham	0.44699	0.778321	0.685958	0.611526	
Haringey	0.341032	0.703754	0.515945	0.458644	
Harrow	0.164108	0.39444	0.177389	0.248781	0.205969
Havering	0.210682	0.536701	0.264977	0.356677	
Hillingdon	0.269062	0.591649	0.321378	1	
Hounslow	0.288107	0.624194	0.373654	0.568464	



Fuzzy kontext

- objekty a atribúty ostali v klasických množinách
- z relácie máme fuzzy množinu $r : B \times A \rightarrow [0, 1]$
- Krajčí: „*Sociálne siete a fuzzy formálna konceptová analýza*”
- $\nearrow : P(B) \rightarrow F(A)$ a $\swarrow : F(A) \rightarrow P(B)$

-

$$\nearrow (X)(a) = \min_{b \in X} \{r(b, a)\}$$

-

$$\swarrow (g) = \{b \in B \mid (\forall a \in A) g(a) \leq r(b, a)\}$$

- Jednostranný fuzzy koncept (X, g) , že

$$\nearrow (X) = g \quad \text{a} \quad \nearrow (g) = X$$



Príklad

	M	F	I	Ch	B
Anna	•	•	•		
Boris	•		•		•
Cyril				•	•
Dana		•		•	•

	M	F	I	Ch	B
Anna	1	0.8	0.8	0.4	0
Boris	0.8	0.6	1	0.4	1
Cyril	0.6	0.4	0.2	1	0.8
Dana	0.6	1	0.4	1	0.8

Príklad

	M	F	I	Ch	B
Anna	1	0.8	0.8	0.4	0
Boris	0.8	0.6	1	0.4	1
Cyril	0.6	0.4	0.2	1	0.8
Dana	0.6	1	0.4	1	0.8

	M	F	I	Ch	B
Anna	■	■	■	■	
Boris	■	■	■	■	■
Cyril	■	■	□	■	■
Dana	■	■	■	■	■

Príklad

	M	F	I	Ch	B
Anna	■	■	■	■	
Boris	■	■	■	■	■
Cyril	■	■	□	■	■
Dana	■	■	■	■	■

{Boris, Dana}; {M/0.6; F/0.6; I/0.4; Ch/0.4; B/0.8}

	M	F	I	Ch	B
Anna					
Boris	■	■	■	■	■
Cyril					
Dana	■	■	■	■	■

Príklad

	M	F	I	Ch	B
Anna	■	■	■	■	
Boris	■	■	■	■	■
Cyril	■	■	□	■	■
Dana	■	■	■	■	■

$\{Anna, Boris, Cyril, Dana\}; \{M/0.6; F/0.4; I/0.2; Ch/0.4; B/0\}$

	M	F	I	Ch	B
Anna	■	■	□	■	
Boris	■	■	□	■	
Cyril	■	■	□	■	
Dana	■	■	□	■	

Príklad

	M	F	I	Ch	B
Anna	■	■	■	■	
Boris	■	■	■	■	■
Cyril	■	■	□	■	■
Dana	■	■	■	■	■

{Boris, Cyril, Dana}; {M/0.6; F/0.4; I/0.2; Ch/0.4; B/0.8}

	M	F	I	Ch	B
Anna					
Boris	■	■	□	■	■
Cyril	■	■	□	■	■
Dana	■	■	□	■	■

Vzdialenosť medzi konceptami

klasická:

$$\rho(X_1, X_2) = 1 - \frac{|X_1 \cap X_2|}{|X_1 \cup X_2|}$$

fuzzy:

$$\rho(X_1, X_2) = 1 - \frac{\sum_{a \in A} \min\{\nearrow (X_1)(a), \nearrow (X_2)(a)\}}{\sum_{a \in A} \max\{\nearrow (X_1)(a), \nearrow (X_2)(a)\}}$$

Rice-Siff algorithm

- 1 $\mathcal{C} \leftarrow \mathcal{D} \leftarrow \{\downarrow\uparrow(\{b\}) \mid b \in B\}$
- 2 while $|\mathcal{D}| > 1$ do
- 3 {
 - 1 $m \leftarrow \min\{\rho(X_1, X_2) \mid X_1, X_2 \in \mathcal{D}, X_1 \neq X_2\}$
 - 2 $\mathcal{E} \leftarrow \{(X_1, X_2) \mid \rho(X_1, X_2) = m\}$
 - 3 $\mathcal{V} \leftarrow \{X \in \mathcal{D} \mid (\exists Y \in \mathcal{D})(X, Y) \in \mathcal{E}\}$
 - 4 $\mathcal{N} \leftarrow \{\downarrow\uparrow(X_1 \cup X_2) \mid (X_1, X_2) \in \mathcal{E}\}$
 - 5 $\mathcal{D} \leftarrow (\mathcal{D} \setminus \mathcal{V}) \cup \mathcal{N}$
 - 6 $\mathcal{C} \leftarrow (\mathcal{C} \cup \mathcal{N})$}
- 4 }
- 5 return $\mathcal{C}$



Crimes of London

- Barnet, Camden , Lambeth , Southwark, Westminster
- VaP/0.180707; TaH/0.38302; D/0.154131; O/0.241968; R/0.224253; CD/0.548116; B/0.670614
-
- Camden, Hackney, Islington, Kensington and Chelsea, Redbridge, Southwark, Westminster
- VaP/0.45971; TaH/0.484228; D/0.340465; O/0.305672; R/0.31612; CD/0.562936; B/0.602967
-
- Brent, Camden, Hackney, Hammersmith and Fulham, Islington, Lambeth, Newham, Southwark, Tower Hamlets, Westminster
- VaP/0.16416; TaH/0.622594; D/0.278537; O/0.38367; R/0.32344; CD/0.685429; B/0.49207

Ďakujem za pozornosť



Jesenná 5, Košice



ics.science.upjs.sk



csl.science.upjs.sk



ui@upjs.sk