

Tato prezentace je součástí wiki-prezentace [Metoda GUHA a systém LISp-Miner](#)

Je dostupná z [této adresy](#)

Verse 30. 7. 2019

Zadávání jednoduchých frekvenčních KL-kvantifikátorů pro proceduru KL-Miner

Jan Rauch

Katedra informačního a znalostního inženýrství

Vysoká škola ekonomická v Praze

Start tlačítkem Add simple frequencies quantifier

KL List of quantifiers settings

Task: Zadávání KL-Miner

Type	Source	Range	Rel.	Value	Units

Example attribute for row: DHodnoceni (3 categories)

Example attribute for column: DStrava_edc5 (5 categories)

	<0,19>	<20,39>	<40,59>	<60,79>	<80,100>
nespokojen	154	400	16	0	0
průměr	0	316	372	262	0
spokojen	0	0	0	124	356

Buttons: Detail, Del, Up, Down, Copy, Paste, Add statistical quantifier, Add simple frequencies quantifier, Add pattern-difference quantifier, Close

- Definice kvantifikátoru se provádí zadáním parametrů 1 - 5.
- Pro všechny parametry se nabízejí defaultní hodnoty.
- Podmínka kvantifikátoru je definována parametry 4 a 5.
- Podmínka kvantifikátoru se aplikuje na hodnotu *IM* spočtenou dle parametru 3.

KL Simple frequencies quantifier settings

1 Category Range

	From	To
Rows	0	100
Columns	0	100

2 Source frequencies

Absolute number

3 Interest measure type

Sum of frequencies

Sum of frequencies from given part of contingency table

4 Relation

Greater than or equal

5 Threshold value

0.5

Threshold-value units

Relative [%] to act condition

Relative to number of rows matching condition. Threshold value [%] is multiplied by (number of rows matching condition)/100

Primary IM Settings

Set as primary IM

From	To
0	1

Normalize value range:

Note

OK Cancel

2

Výchozí KL-tabulka

KL-vztah $R \approx C/\chi$ s jednoduchým frekvnečním kvantifikátorem se verifikuje na základě KL-tabulky $KL(R,C,M/\chi)$. Využije se i KL-tabulka $KL(R,C,M)$.

$$KL(R,C,M/\chi) =$$

M/χ	c_1	\dots	c_u	\dots	c_v	\dots	c_L	Σ_l
r_1	$n_{1,1}$	\dots	$n_{1,u}$	\dots	$n_{1,v}$	\dots	$n_{1,L}$	$n_{1,*}$
\vdots	\vdots	\ddots	\vdots	\ddots	\vdots	\ddots	\vdots	\vdots
r_p	$n_{p,1}$	\dots	$n_{p,u}$	\dots	$n_{p,v}$	\dots	$n_{p,L}$	$n_{p,*}$
\vdots	\vdots	\ddots	\vdots	\ddots	\vdots	\ddots	\vdots	\vdots
r_q	$n_{q,1}$	\dots	$n_{q,u}$	\dots	$n_{q,v}$	\dots	$n_{q,L}$	$n_{q,*}$
\vdots	\vdots	\ddots	\vdots	\ddots	\vdots	\ddots	\vdots	\vdots
r_K	$n_{K,1}$	\dots	$n_{K,u}$	\dots	$n_{K,v}$	\dots	$n_{K,L}$	$n_{K,*}$
Σ_k	$n_{*,1}$	\dots	$n_{*,u}$	\dots	$n_{*,v}$	\dots	$n_{*,L}$	n

$$KL(R,C,M) =$$

M	c_1	\dots	c_L	Σ_l
r_1	$m_{1,1}$	\dots	$m_{1,L}$	$m_{1,*}$
\vdots	\vdots		\vdots	\vdots
r_K	$m_{K,1}$	\dots	$m_{K,L}$	$m_{K,*}$
Σ_k	$m_{*,1}$	\dots	$m_{*,L}$	n_T

Pokud $\chi = True$, pak se KL-vztah týká celé matice dat a platí $n_T = n$.

1 - Category Range

- Category Range určuje výsek KL-tabulky $KL(R,C,M/\chi)$, na který budou aplikovány ostatní parametry.
- Provádí se způsobem popsáním [zde](#).
- Výsek KL-tabulky $KL(R,C,M/\chi)$ je dán čtveřicí $\langle p,q,u,v \rangle$ celých čísel splňujících $1 \leq p \leq q \leq K$ a $1 \leq u \leq v \leq L$
- Výsek $\langle p,q,r,s \rangle$ $KL(R,C,M/\chi)$ zahrnuje frekvence pro kategorie r_p, \dots, r_q a c_u, \dots, c_v .

Category Range

	From	To	
Rows	p	q	Reset to All
Columns	u	v	

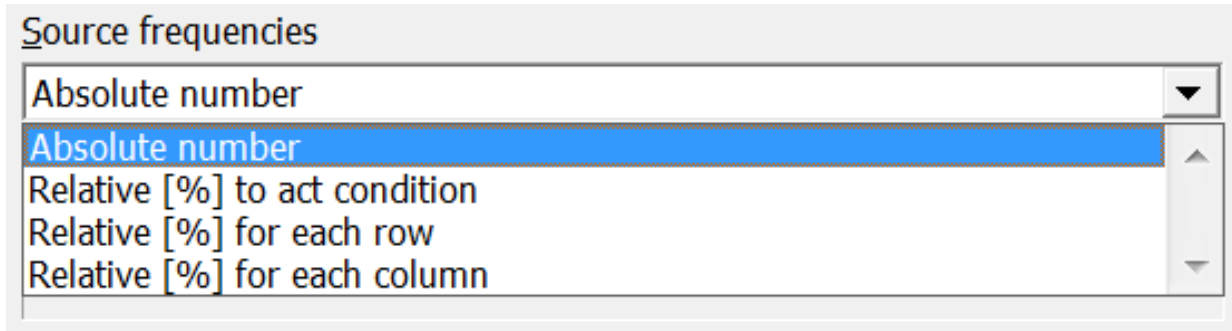
Absolute category index

Relative range [%] to act number of categories

M/χ	c_1	...	c_u	...	c_v	...	c_L	Σ_l
r_1	$n_{1,1}$...	$n_{1,u}$...	$n_{1,v}$...	$n_{1,L}$	$n_{1,*}$
\vdots	\vdots	\ddots	\vdots	\ddots	\vdots	\ddots	\vdots	\vdots
r_p	$n_{p,1}$...	$n_{p,u}$...	$n_{p,v}$...	$n_{p,L}$	$n_{p,*}$
\vdots	\vdots	\ddots	\vdots	\ddots	\vdots	\ddots	\vdots	\vdots
r_q	$n_{q,1}$...	$n_{q,u}$...	$n_{q,v}$...	$n_{q,L}$	$n_{q,*}$
\vdots	\vdots	\ddots	\vdots	\ddots	\vdots	\ddots	\vdots	\vdots
r_K	$n_{K,1}$...	$n_{K,u}$...	$n_{K,v}$...	$n_{K,L}$	$n_{K,*}$
Σ_k	$n_{*,1}$...	$n_{*,u}$...	$n_{*,v}$...	$n_{*,L}$	n

2 - Source frequencies

Vybírá se jedna z voleb nabízených v menu *Source frequencies*.

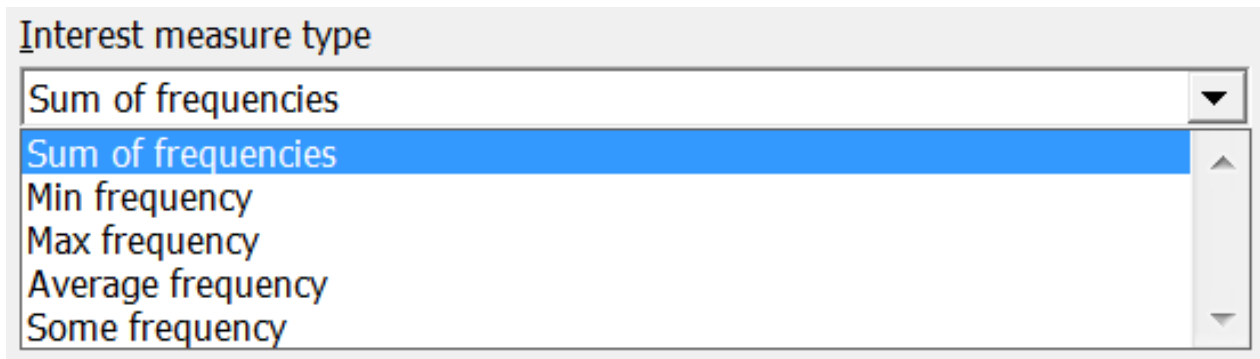


Na základě této volby se určí podmatice $\{z_{i,j}\}_{i=1,\dots,p}^{j=u,\dots,v}$ frekvencí pro výpočet hodnoty IM . Platí:

- pro *Absolute number* $\{z_{i,j}\}_{i=p,\dots,q}^{j=u,\dots,v} = \{n_{i,j}\}_{i=p,\dots,q}^{j=u,\dots,v}$
- pro *Relative [%] to act condition* $\{z_{i,j}\}_{i=p,\dots,q}^{j=u,\dots,v} = \{n_{i,j}/n\}_{i=p,\dots,q}^{j=u,\dots,v}$
- pro *Relative [%] for each row* $\{z_{i,j}\}_{i=p,\dots,q}^{j=u,\dots,v} = \{n_{i,j}/n_{i,*}\}_{i=p,\dots,q}^{j=u,\dots,v}$
- pro *Relative [%] for each column* $\{z_{i,j}\}_{i=p,\dots,q}^{j=u,\dots,v} = \{n_{i,j}/n_{*,j}\}_{i=p,\dots,q}^{j=u,\dots,v}$

3 - Inteterest measure type

Vybírá se jedna z voleb nabízených v menu *Interest measure type*.

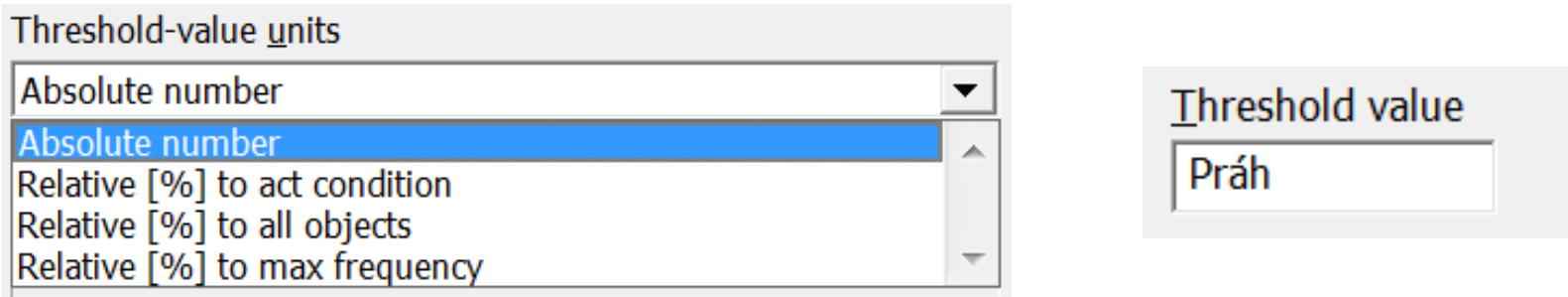


Na základě této volby se určí hodnota IM pro porovnání s prahem dle relace zadané parametrem 5. Použije se podmatice $\{z_{i,j}\}_{i=1,\dots,p}^{j=u,\dots,v}$. Platí:

- pro *Sum of frequencies* se spočítá hodnota $IM = \sum_{i=p}^q \sum_{j=u}^v z_{i,j}$
- pro *Min frequency* se spočítá hodnota $IM = \min_{i,j} \{z_{i,j} | i = p, \dots, q; j = u, \dots, v\}$
- pro *Max frequency* se spočítá hodnota $IM = \max_{i,j} \{z_{i,j} | i = p, \dots, q; j = u, \dots, v\}$
- pro *Average frequency* se spočítá $IM = \frac{\sum_{i=p}^q \sum_{j=u}^v z_{i,j}}{((q-p+1)(v-u+1))}$
- pro *Some frequency* se zjišťuje, zda alespoň jedno $z_{i,j}$ pro $i = p, \dots, q; j = u, \dots, v$ splňuje zadanou nerovnost.

4 - Threshold-value units

Vybírá se jedna z voleb nabízených v menu *Threshold-value units*



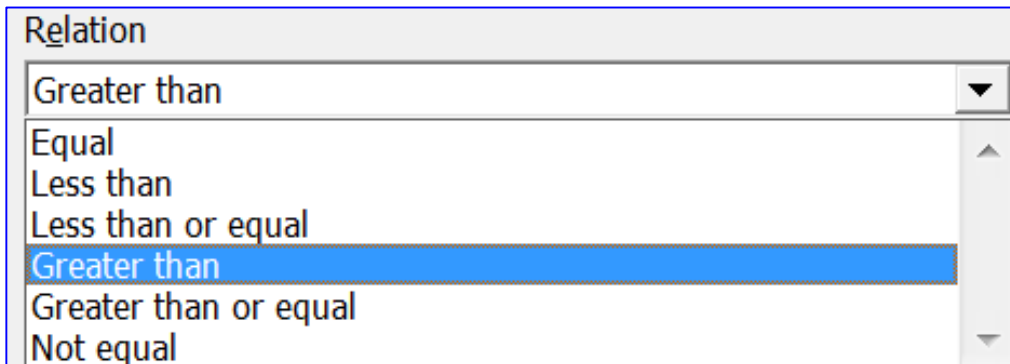
The image shows a software interface with two main components. On the left is a dropdown menu titled "Threshold-value units". The menu is open, showing four options: "Absolute number" (which is highlighted in blue), "Relative [%] to act condition", "Relative [%] to all objects", and "Relative [%] to max frequency". On the right is a text input field titled "Threshold value" containing the word "Práh".

Na základě této volby se určí jednotky, ve kterých je hodnota *Práh* zadaná v poli *Threshold value*. To znamená, že se před porovnáním dle parametru 5 zadaná hodnota *Práh* vynásobí vhodným číslem dle volby v menu. Výslednou hodnotu pro porovnání značíme *PráhAkt*. Platí:

- pro *Absolute number* $PráhAkt = Práh$
- pro *Relative [%] to act condition* $PráhAkt = \frac{100}{n} * Práh$
- pro *Relative [%] to all objects* $PráhAkt = \frac{100}{n_T} * Práh$
- pro *Relative [%] to max frequency* $PráhAkt = \frac{100}{Max} * Práh$ kde $Max = \max_{i,j} \{ m_{i,j} | i = 1, \dots, K; j = 1, \dots, L \}$

5 - Relation x Threshold value

Na základě volby v nabídce *Relation* se vybere relace, která se použije pro porovnání hodnoty *IM* vypočtené dle parametru 3 - *Inte*nterest measure type s hodnotou *PráhAkt* vypočtenou dle parametru 4 - *Threshold-value units*. Pro volbu *Some frequency* u parametru 3 musí být tato relace platná alespoň jedno $z_{i,j}$ z podmatice $\{z_{i,j}\}_{i=1,\dots,p}^{j=u,\dots,v}$.



Relation

Greater than

Equal

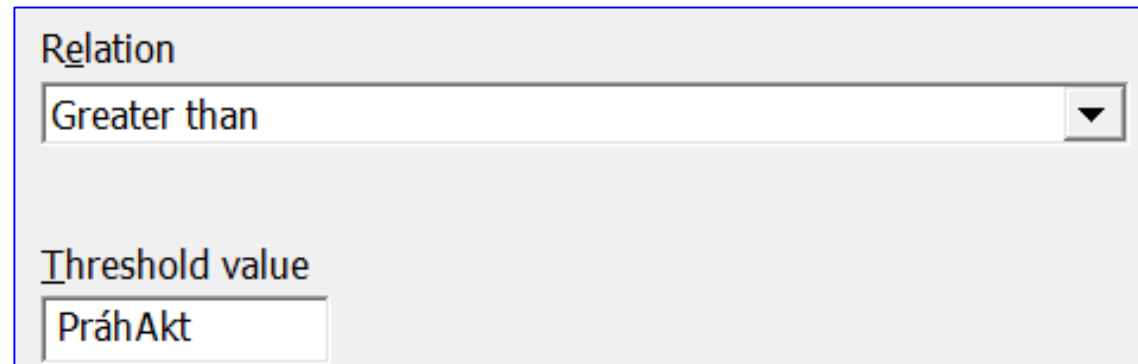
Less than

Less than or equal

Greater than

Greater than or equal

Not equal



Relation

Greater than

Threshold value

PráhAkt

Platnost vybrané relace je považována za podmínku definující KL-kvantifikátor zadaný parametry 1 až 5.