

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

Je dostupná z [této adresy](#)

Verze 28. 7. 2019

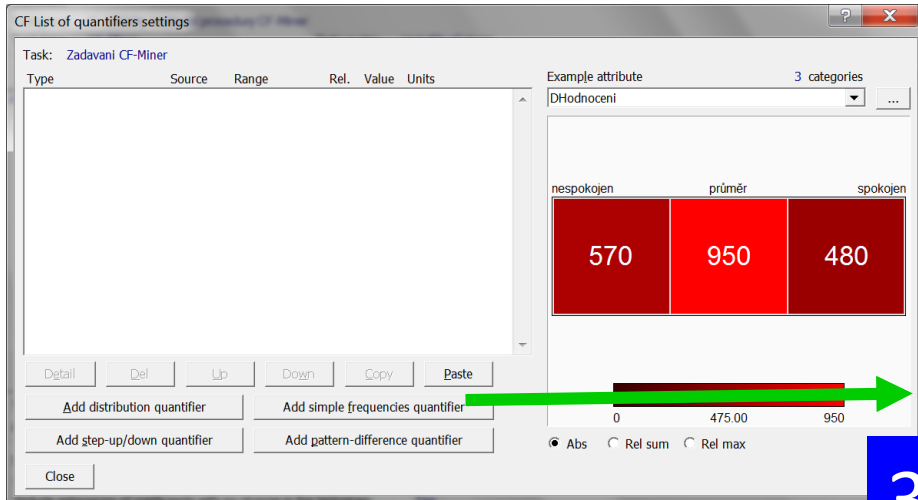
Definice jednoduchých frekvenčních CF-kvantifikátorů pro proceduru CF-Miner

Jan Rauch

Katedra informačního a znalostního inženýrství

Vysoká škola ekonomická v Praze

Start tlačítkem Add simple frequencies quantifier



- Definice se provádí zadáním parametrů 1 - 5.
- Pro všechny parametry se nabízejí defaultní hodnoty.
- Podmínka kvantifikátoru je definována parametrem 5.

3 Interest measure type: Some frequency
At least one frequency from range of categories

5 Relation: Greater than or equal

Threshold value: 50

2 Source frequencies: Absolute number
Absolute frequencies. Frequencies in histogram are left as given.

4 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

1 Category Range: From 0 To 100
Reset to All

Absolute category index
 Relative range [%] to act number of categories

Primary IM Settings:
 Set as primary IM
 Normalize value range: From 0 To 1

Note: -

OK Cancel

Výchozí CF-tabulka

Jednoduchý frekvenční CF-kvantifikátor se pro CF-výraz $\approx A/\chi$ verifikuje na základě [rozšířené CF-tabulky CF\(A, \$\chi\$, M\)](#):

četnosti kategorií pro atribut A	a_1	...	a_p	...	a_q	...	a_K	Σ
absolutní četnosti v matici M/ χ	n_1	...	n_p	...	n_q	...	n_K	n
relativní četnosti v matici M/ χ	$f_1=n_1/n$...	$f_p=n_p/n$...	$f_q=n_q/n$...	$f_K=n_K/n$	1
absolutní četnosti v matici M	m_1	...	m_p	...	m_q	...	m_K	n_T
relativní četnosti mezi M/ χ a M	$v_1= n_1/m_1$...	$v_p= n_p/m_p$...	$v_q= n_q/m_q$...	$v_K= n_K/m_K$	-

Výraz $\approx A$ týkající se rozložení histogramu na celé matici dat chápeme jako $\approx A/True$.

1 - Category Range

- Category Range určuje výsek CF-tabulky, na který budou aplikovány ostatní parametry.
- Provádí se způsobem popsáním [zde](#).
- Výsek CF-tabulky týkající se K kategorií je dvojice $\langle p, q \rangle$ celých čísel splňujících $1 \leq p \leq q \leq K$.
- Výsek $\langle p, q \rangle$ CF-tabulky zahrnuje frekvence pro kategorie a_p, \dots, a_q .

Category Range

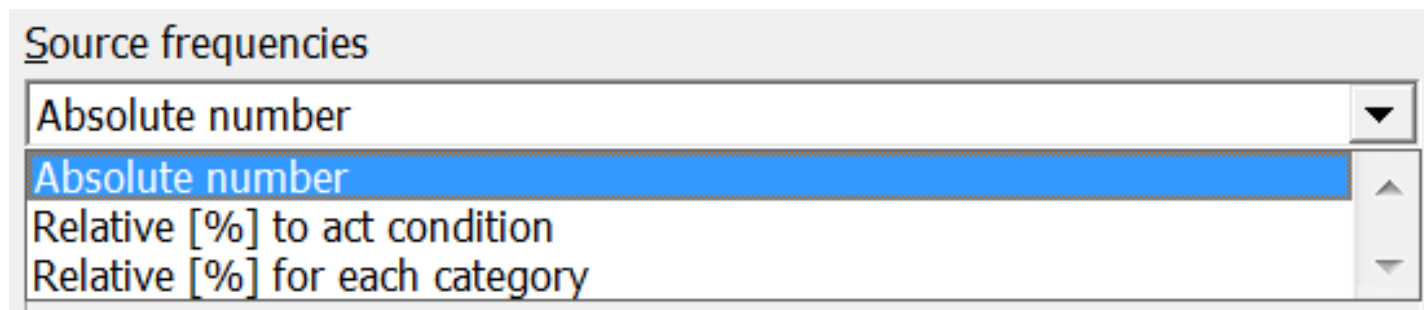
From To Reset to All

Absolute category index
 Relative range [%] to act number of categories

četnosti kategorií pro atribut A	a_1	...	a_p	...	a_q	...	a_K	Σ
absolutní četnosti v matici M/χ	n_1	...	n_p	...	n_q	...	n_K	n
relativní četnosti v matici M/χ	$f_1 = n_1/n$...	$f_p = n_p/n$...	$f_q = n_q/n$...	$f_K = n_K/n$	1
absolutní četnosti v matici M	m_1	...	m_p	...	m_q	...	m_K	n_T
relativní četnosti mezi M/χ a M	$v_1 = n_1/m_1$...	$v_p = n_p/m_p$...	$v_q = n_q/m_q$...	$v_K = n_K/m_K$	-

2 - Source frequencies

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

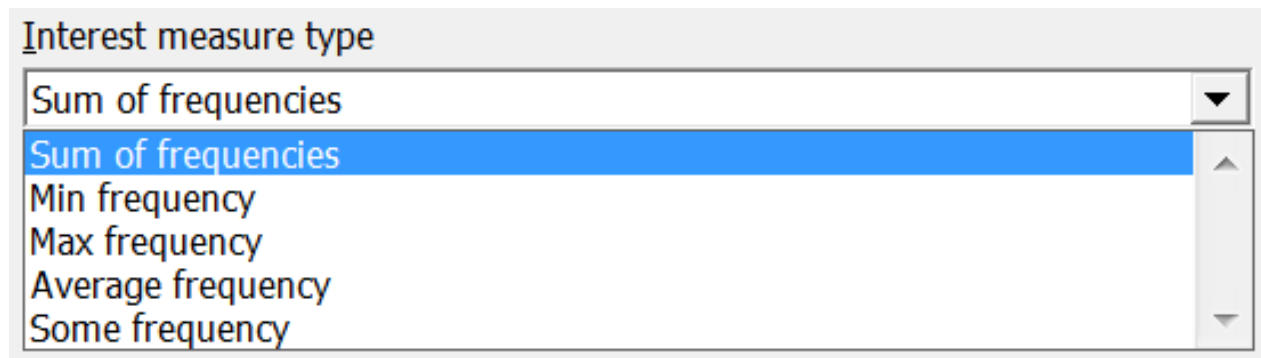


Na základě této volby se určí vektor $\langle w_p, \dots, w_q \rangle$ frekvencí použitý při verifikaci. Platí:

- pro *Absolute number* $\langle w_p, \dots, w_q \rangle = \langle n_p, \dots, n_q \rangle$, jedná se tedy o [histogram absolutních frekvencí](#)
- pro *Relative [%] to act condition* $\langle w_p, \dots, w_q \rangle = \langle f_p, \dots, f_q \rangle$, jedná se tedy o [histogram podílu kategorií na podmínce](#)
- pro *Relative [%] for each category* $\langle w_p, \dots, w_q \rangle = \langle v_p, \dots, v_q \rangle$, jedná se tedy o [histogram podílu podmínky na kategoriích](#)

3 - Interest measure type

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

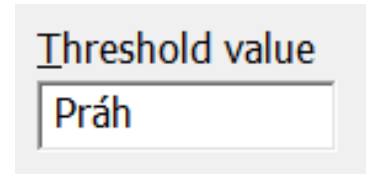
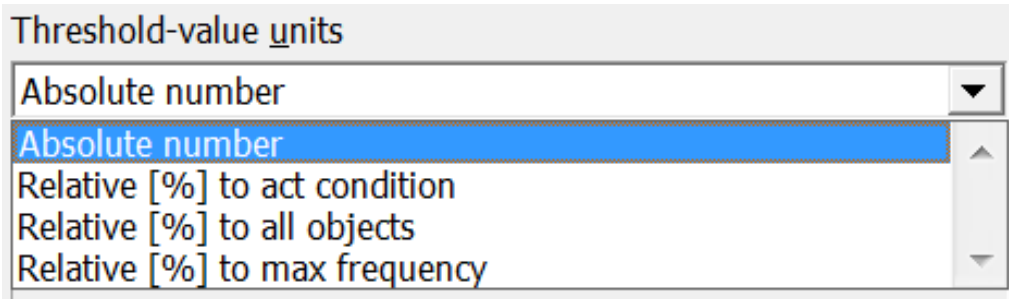


Na základě této volby se určí hodnota IM pro porovnání s prahem dle relace zadané parametrem 5. Použije se vektor $\langle w_p, \dots, w_q \rangle$. Platí:

- pro *Sum of frequencies* se spočítá hodnota $IM = \sum_{i=p}^q w_i$
- pro *Min frequency* se spočítá hodnota $IM = \min_i \{w_i | i = p, \dots, q\}$
- pro *Max frequency* se spočítá hodnota $IM = \max_i \{w_i | i = p, \dots, q\}$
- pro *Average frequency* se spočítá $IM = \frac{\sum_{i=p}^q w_i}{(q-p+1)}$
- pro *Some frequency* se zjišťuje, zda alespoň jedno w_i pro $i = p, \dots, q$ splňuje nerovnost zadanou parametrem 5

4 - Threshold-value units

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

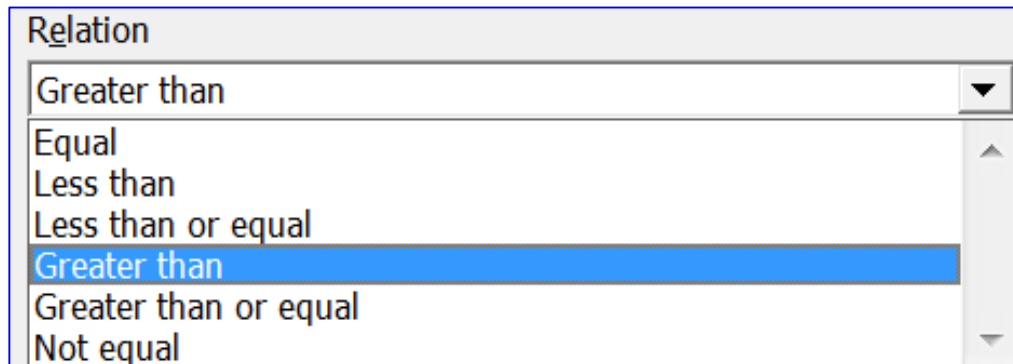


Na základě této volby se určí jednotky, ve kterých je hodnota *Práh* zadaná v poli *Threshold value*. To znamená, že se před porovnáním dle parametru 5 zadaná hodnota *Práh* vynásobí vhodným číslem dle volby v menu. Výslednou hodnotu pro porovnání značíme *PráhAkt*. Platí:

- 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 $Max = \max_i \{n_i | i = 1, \dots, K\}$

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 w_i z vektoru $\langle w_p, \dots, w_q \rangle$.



Relation

Greater than

Equal

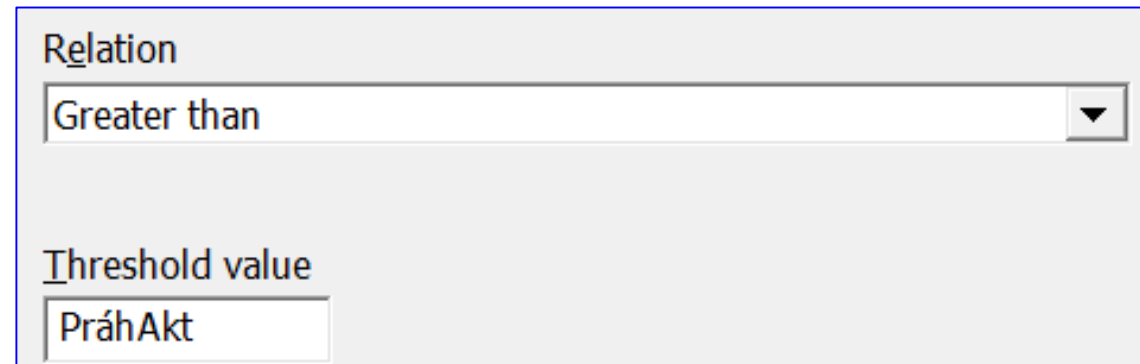
Less than

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 CF-kvantifikátor zadaný parametry 1 až 5.