

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

Je dostupná z [této adresy](#)

Verze 1. 8. 2019

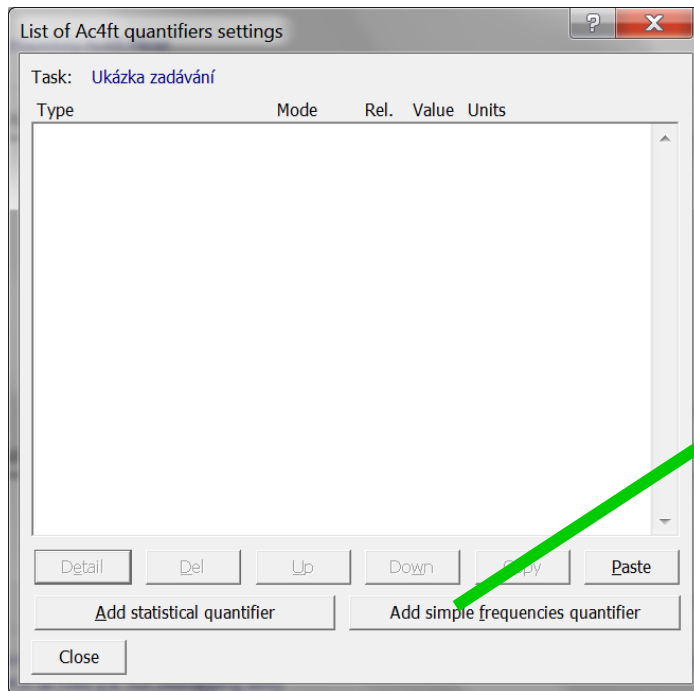
Zadávání jednoduchých frekvenčních Ac4ft-kvantifikátorů pro proceduru Ac4ft-Miner

Jan Rauch

Katedra informačního a znalostního inženýrství

Vysoká škola ekonomická v Praze

Start tlačítkem Add simple frequencies quantifier



- Definice se provádí zadáním parametrů 1 - 4.
- Pro všechny parametry se nabízejí defaultní hodnoty.
- Podmínka kvantifikátoru je definována parametrem 4.

Ac4ft Simple frequencies quantifier settings

Interest measure type

Sum of frequencies

Sum of frequencies from source contingency table

Relation

Greater than or equal

Threshold value

0

Operation mode

Delta of absolute frequencies values

Quantifier test applied to the 4ft-table computed by subtracting of corresponding frequencies

Threshold-value units

Absolute number

Absolute number. Threshold value is left as given.

Primary IM Settings

Set as primary IM

From To

Normalize value range: 0 1

Note

OK Cancel

1 2 3 4

Výchozí čtyřpolní tabulky

Statistický Ac4ft-kvantifikátor se aplikuje na Ac4ft-tabulku $Ac4ft(\phi \wedge \Phi \approx \psi \wedge \Psi, M)$ pro akční pravidlo $\phi \wedge \Phi \approx \psi \wedge \Psi$ tvořenou čtyřpolními tabulkami T_{Init} a T_{Fin}

$$T_{Init} =$$

M	Ψ_{Init}	$\neg\Psi_{Init}$
Φ_{Init}	a_I	b_I
$\neg\Phi_{Init}$	c_I	d_I

$$T_{Fin} =$$

M	Ψ_{Fin}	$\neg\Psi_{Fin}$
Φ_{Fin}	a_F	b_F
$\neg\Phi_{Fin}$	c_F	d_F

Počet řádků v matici dat M značíme n_T .

Pokud vyhodnocujeme podmíněné Ac4ft-pravidlo $\alpha \times \beta : \phi \sim \psi / \chi$, pak platí:

$$T_{Init} =$$

M/ χ	Ψ_{Init}	$\neg\Psi_{Init}$
Φ_{Init}	a_I	b_I
$\neg\Phi_{Init}$	c_I	d_I

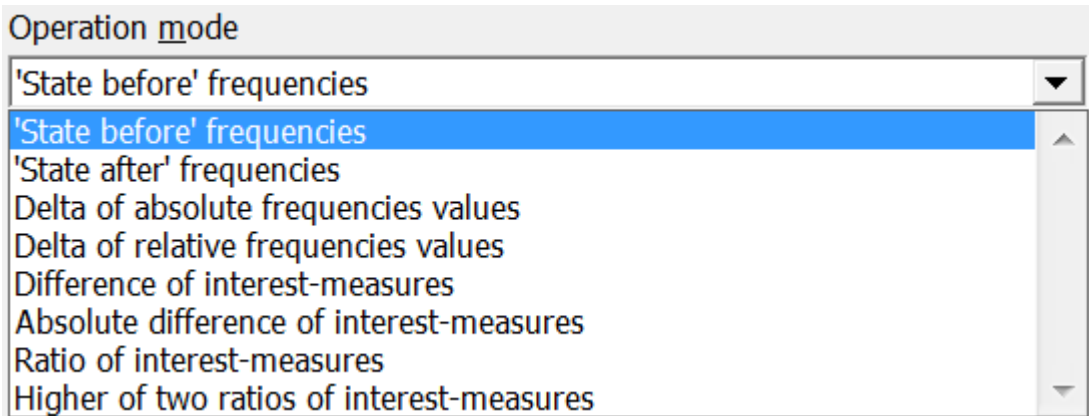
$$T_{Fin} =$$

M/ χ	Ψ_{Fin}	$\neg\Psi_{Fin}$
Φ_{Fin}	a_F	b_F
$\neg\Phi_{Fin}$	c_F	d_F

Počet řádků v matici dat M/ χ značíme n . Pokud $\chi = True$, pak $n = n_T$.

1 - Operation mode

Vybírá se jeden z operačních módů nabízených v menu *Operation mode*. Ten určuje, jakým způsobem bude aplikována vybraná míra zajímavosti na čtyřpolní tabulky T_{Init} a T_{Fin} .



Pro zbývající čtyři nabízené módy se vybraná míra zajímavosti aplikuje zvlášť na každou z tabulek T_{Init} a T_{Fin} . Poté se výsledky dále zpracují.

Pro první čtyři nabízené Operation mode se nejprve vytvoří čtyřpolní tabulka $\langle a, b, c, d \rangle$ takto:

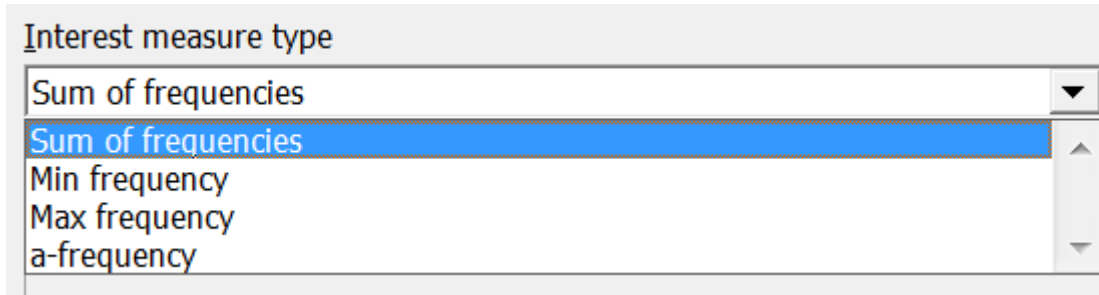
- pro '"State before" frequencies' $\langle a, b, c, d \rangle = \langle a_I, b_I, c_I, d_I \rangle$
- pro '"State after" frequencies' $\langle a, b, c, d \rangle = \langle a_F, b_F, c_F, d_F \rangle$

- pro *Delta of absolute frequencies values*
 $\langle a, b, c, d \rangle = \langle |a_I - a_F|, |b_I - b_F|, |c_I - c_F|, |d_I - d_F| \rangle$

- pro *Delta of relative frequencies values*
 $\langle a, b, c, d \rangle = \langle \left| \frac{a_I}{n_I} - \frac{a_F}{n_F} \right|, \left| \frac{b_I}{n_I} - \frac{b_F}{n_F} \right|, \left| \frac{c_I}{n_I} - \frac{c_F}{n_F} \right|, \left| \frac{d_I}{n_I} - \frac{d_F}{n_F} \right| \rangle$
kde $n_I = a_I + b_I + c_I + d_I$ a $n_F = a_F + b_F + c_F + d_F$.

2 - Interest measure type (A)

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



Pro výpočet hodnoty *IM* pro první čtyři operační módy

- *First set frequencies*
- *Second set frequencies*
- *Delta of absolute frequencies values,*
- *Delta of relative frequencies values*

pro jednotlivé volby *Interest measure type* se použije čtyřpolní tabulka $\langle a, b, c, d \rangle$ přiřazená výše uvedeným způsobem.

Interest measure type	Hodnota <i>IM</i>
Sum of frequencies	$a + b + c + d$
Min frequency	$\min\{a, b, c, d\}$
Max frequency	$\max\{a, b, c, d\}$
a-frequency	a

2 - Interest measure type (B)

Při výpočtu hodnoty IM pro zbývající čtyři operační módy se nejprve se vypočtou hodnoty IM_I a IM_F pro vybraný *Interest measure type*:

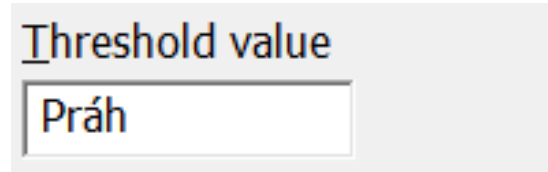
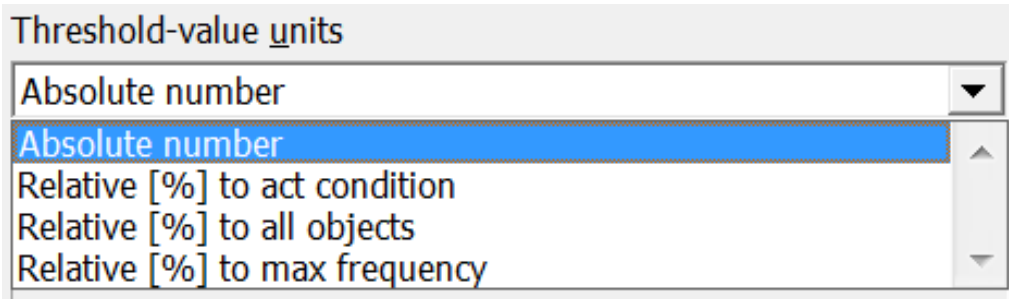
Interest measure type	IM_I	IM_F
Sum of frequencies	$a_I + b_I + c_I + d_I$	$a_F + b_F + c_F + d_F$
Min frequency	$\min\{a_I, b_I, c_I, d_I\}$	$\min\{a_F, b_F, c_F, d_F\}$
Max frequency	$\max\{a_I, b_I, c_I, d_I\}$	$\max\{a_F, b_F, c_F, d_F\}$
a frequency	a_I	a_F

Poté se vypočte hodnota IM v závislosti na vybraném *Operation mode*:

Operation mode	IM
Difference of interest-measures	$IM_I - IM_F$
Absolute difference of interest-measures	$ IM_I - IM_F $
Ratio of interest-measures	$\frac{IM_I}{IM_F}$
Higher of ratios of interest-measures	$\max\left\{\frac{IM_I}{IM_F}, \frac{IM_F}{IM_I}\right\}$

3 - Threshold-value units

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

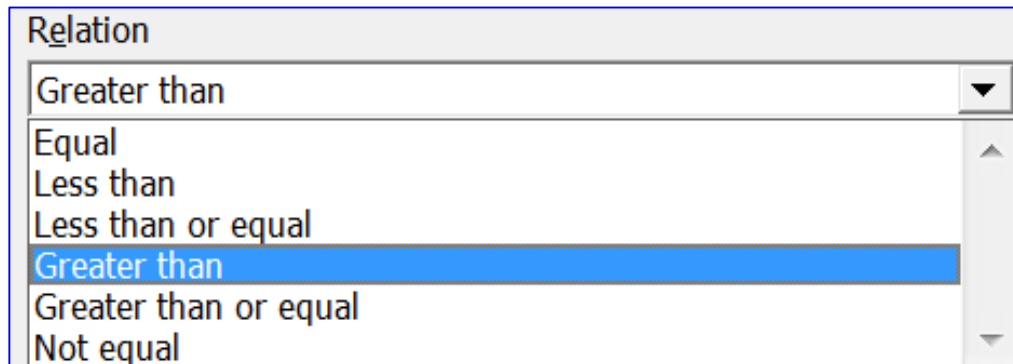


Na základě této volby se určí jednotky, ve kterých je hodnota *Práh* zadaná v poli *Threshold value*. To znamená, že se před porovnáním dle parametru 4 zadaná hodnota *Práh* vynásobí vhodným číslem dle volby v menu. Výslednou hodnotu pro porovnání značíme *PráhAkt*. Platí:

- 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 all objects* $PráhAkt = \frac{100}{Max} * Práh$ kde $Max = \max\{a, b, c, d\}$

4 - Relation x Threshold value

Na základě volby v nabídce *Relation* se vybere relace, která se použije pro porovnání hodnoty *IM* vypočtené dle parametru 2 - *Inte*nterest measure type s hodnotou *PráhAkt* vypočtenou dle parametru 4 - *Threshold-value units*.



Relation

Greater than

Equal

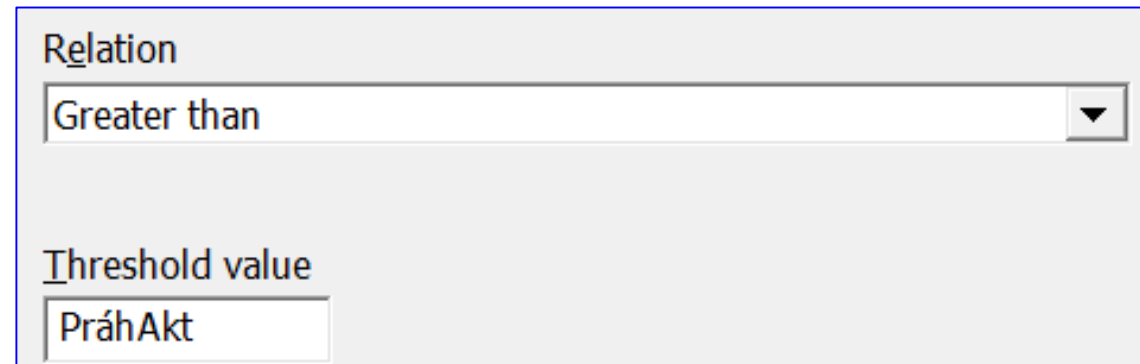
Less than

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 Ac4ft-kvantifikátor zadaný parametry 1 až 4.