

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

Je dostupná z [této adresy](#)

Verse 24. 7. 2019

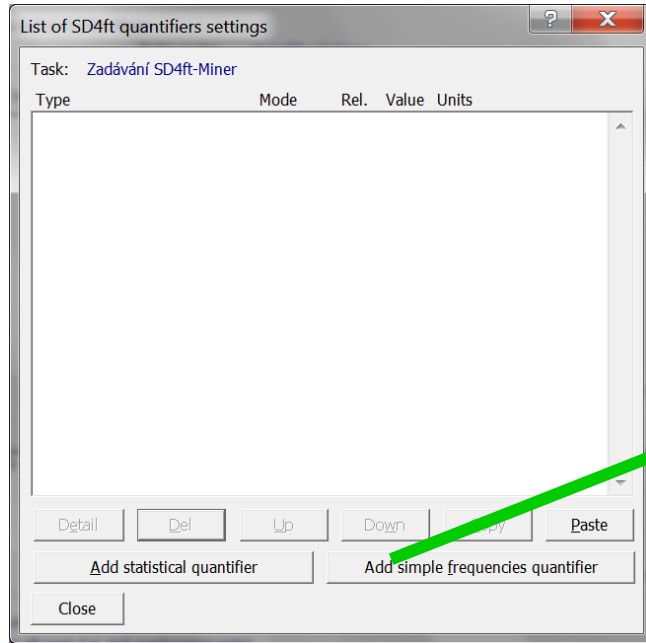
Zadávání jednoduchých frekvenčních SD4ft-kvantifikátorů pro proceduru SD4ft-Miner

Jan Rauch

Katedra informačního a znalostního inženýrství

Vysoká škola ekonomická v Praze

Start tlačítkem Add simple frequencies quantifier



Zadání kvantifikátoru se provádí ve čtyřech krocích.

2 Interest measure type
Sum of frequencies
Sum of frequencies from source contingency table

4 Relation
Greater than or equal

3 Threshold value units
Absolute number
Absolute number. Threshold value is left as given.

1 Operation mode
Delta of absolute frequencies values
Quantifier test applied to the frequency table computed by subtracting corresponding frequencies

Threshold value
0

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

Note
-

OK Cancel

Výchozí čtyřpolní tabulky

SD4ft-kvantifikátor se aplikuje na čtyřpolní tabulky T_α a T_β pro vyhodnocované SD4ft-pravidlo.

Pokud vyhodnocujeme nepodmíněné SD4ft-pravidlo $\alpha \times \beta : \varphi \sim \psi$, pak platí:

$$T_\alpha =$$

M/ α	ψ	$\neg\psi$
φ	a_α	b_α
$\neg\varphi$	c_α	d_α

$$T_\beta =$$

M/ β	ψ	$\neg\psi$
φ	a_β	b_β
$\neg\varphi$	c_β	d_β

Pokud vyhodnocujeme nepodmíněné SD4ft-pravidlo $\alpha \times \beta : \varphi \sim \psi / \chi$, pak platí:

$$T_\alpha =$$

M/ $\alpha \wedge \chi$	ψ	$\neg\psi$
φ	a_α	b_α
$\neg\varphi$	c_α	d_α

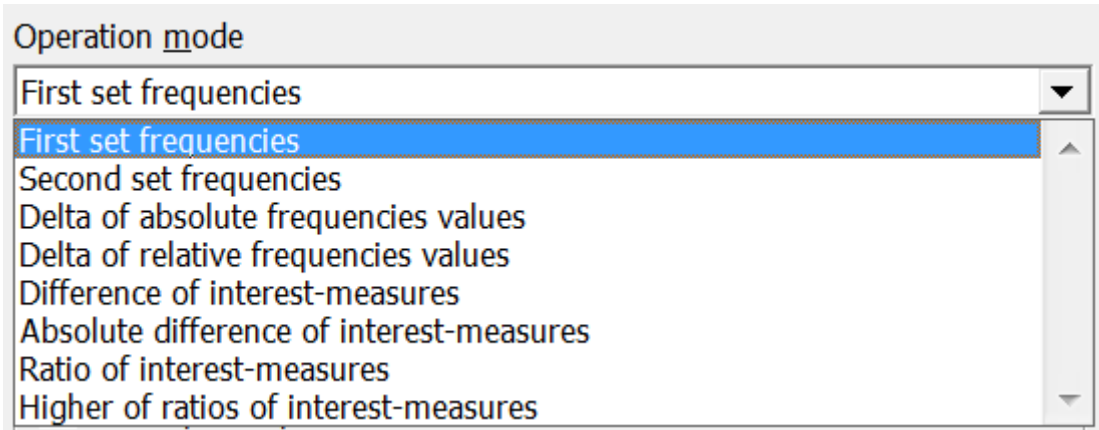
$$T_\beta =$$

M/ $\beta \wedge \chi$	ψ	$\neg\psi$
φ	a_β	b_β
$\neg\varphi$	c_β	d_β

Dalším východiskem pro vyhodnocení daného SD4ft-pravidla je počet řádků analyzované matice dat M dále značený n_T .

Krok 1 - výběr operačního módu

Vybírá se jeden z operačních módů nabízených v menu *Operation mode*. Ten určuje, jakým způsobem bude aplikována vybraná míra zajímavosti na čtyřpolní tabulky T_α a T_β .



Pro zbývající čtyři nabízené módy se vybraná míra zajímavosti aplikuje zvlášť na každou z tabulek T_α a T_β . Poté se výsledky dále zpracují.

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

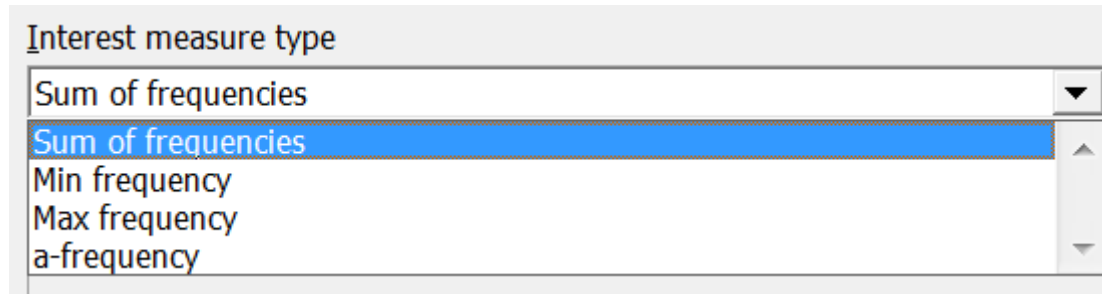
- pro *First set frequencies* $\langle a, b, c, d \rangle = \langle a_\alpha, b_\alpha, c_\alpha, d_\alpha \rangle$
- pro *Second set frequencies* $\langle a, b, c, d \rangle = \langle a_\beta, b_\beta, c_\beta, d_\beta \rangle$
- pro *Delta of absolute frequencies values*
 $\langle a, b, c, d \rangle = \langle |a_\alpha - a_\beta|, |b_\alpha - b_\beta|, |c_\alpha - c_\beta|, |d_\alpha - d_\beta| \rangle$
- pro *Delta of relative frequencies values*

$$\langle a, b, c, d \rangle = \left\langle \left| \frac{a_\alpha}{n_\alpha} - \frac{a_\beta}{n_\beta} \right|, \left| \frac{b_\alpha}{n_\alpha} - \frac{b_\beta}{n_\beta} \right|, \left| \frac{c_\alpha}{n_\alpha} - \frac{c_\beta}{n_\beta} \right|, \left| \frac{d_\alpha}{n_\alpha} - \frac{d_\beta}{n_\beta} \right| \right\rangle$$

$$\text{kde } n_\alpha = a_\alpha + b_\alpha + c_\alpha + d_\alpha \text{ a } n_\beta = a_\beta + b_\beta + c_\beta + d_\beta.$$

Krok 2 - výběr míry zajímavosti (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

Krok 2 - výběr míry zajímavosti (B)

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

Interest measure type	IM_α	IM_β
Sum of frequencies	$a_\alpha + b_\alpha + c_\alpha + d_\alpha$	$a_\beta + b_\beta + c_\beta + d_\beta$
Min frequency	$\min\{a_\alpha, b_\alpha, c_\alpha, d_\alpha\}$	$\min\{a_\beta, b_\beta, c_\beta, d_\beta\}$
Max frequency	$\max\{a_\alpha, b_\alpha, c_\alpha, d_\alpha\}$	$\max\{a_\beta, b_\beta, c_\beta, d_\beta\}$
a frequency	a_α	a_β

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

Operation mode	IM
Difference of interest-measures	$IM_\alpha - IM_\beta$
Absolute difference of interest-measures	$ IM_\alpha - IM_\beta $
Ratio of interest-measures	$\frac{IM_\alpha}{IM_\beta}$
Higher of ratios of interest-measures	$\max\left(\frac{IM_\alpha}{IM_\beta}, \frac{IM_\beta}{IM_\alpha}\right)$

Krok 3 - výpočet hodnoty pro porovnání s prahem

Z voleb v nabídce *Threshold-value units* se vybere způsob výpočtu hodnoty $IMTh$, která bude porovnávána s hodnotou zadanou v poli *Threshold value*.

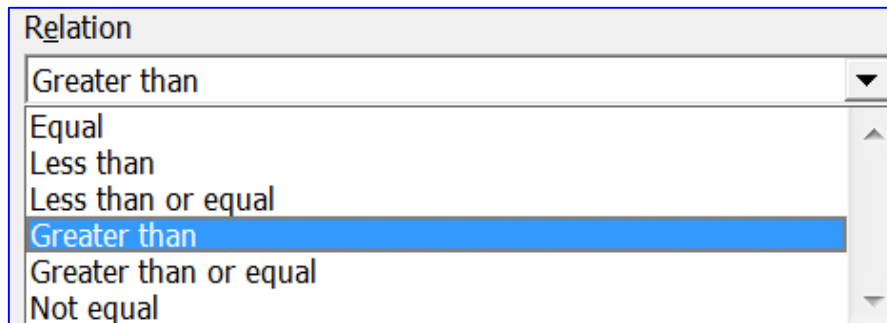
Threshold-value units	nabídka v <i>Threshold-value units</i>	$IMTh$
Absolute number	Absolute number	IM
Absolute number	Relative [%] to act condition	$100 * \frac{IM}{n}$
Relative [%] to act condition	Relative [%] to all objects	$100 * \frac{IM}{n_T}$
Relative [%] to all objects	Relative [%] to max. frequency	$100 * \frac{IM}{Max}$
Relative [%] to max frequency		

Připomeňme, že n_T značí počet řádků analyzované matice dat M .

Dále platí, že Max je maximální frekvence z frekvencí uvedených ve čtyřpolní tabulce $\langle a, b, c, d \rangle$ vytvořené dle kroku 1.

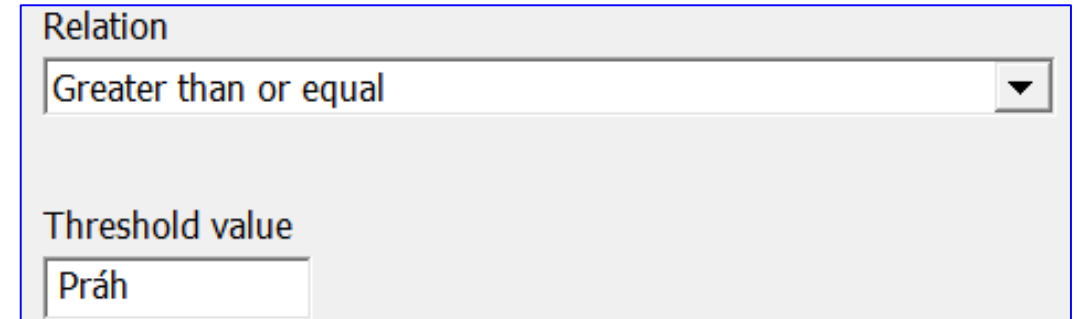
Krok 4 - porovnání vypočtené hodnoty s prahem

Na základě volby v nabídce *Relation* se vybere relace která bude použita pro porovnání vypočtené hodnoty *IMTh* s hodnotou zadanou v poli *Threshold value*.



Relation

- Greater than
- Equal
- Less than
- Less than or equal
- Greater than
- Greater than or equal
- Not equal



Relation

Greater than or equal

Threshold value

Práh

Platnost vybrané relace mezi hodnotou *IMTh* a hodnotou zadanou v poli *Threshold value* je považována za podmínku definující SD4ft-kvantifikátor zadaný v krocích 1 až 4.