

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

Je dostupná z [této adresy](#)

Verze 1. 8. 2019

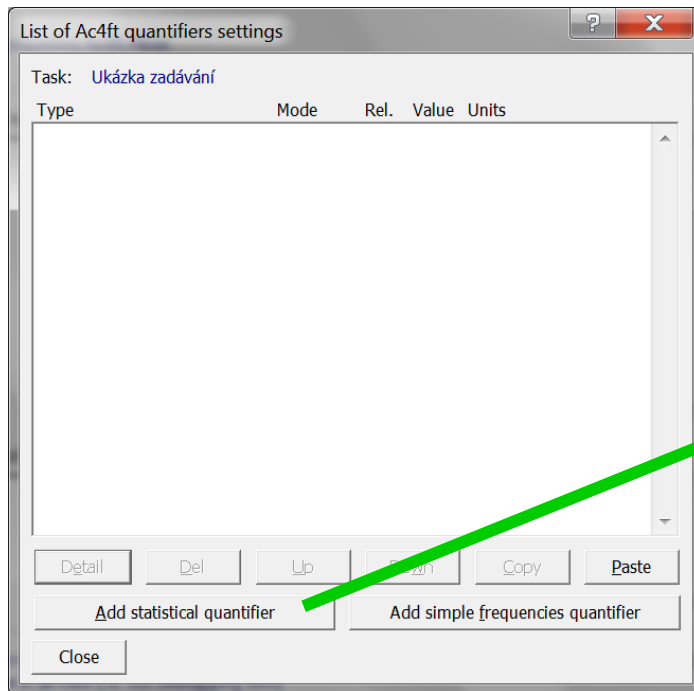
Zadávání statistických Ac4ft-kvantifikátorů pro proceduru Ac4ft-Miner

Jan Rauch

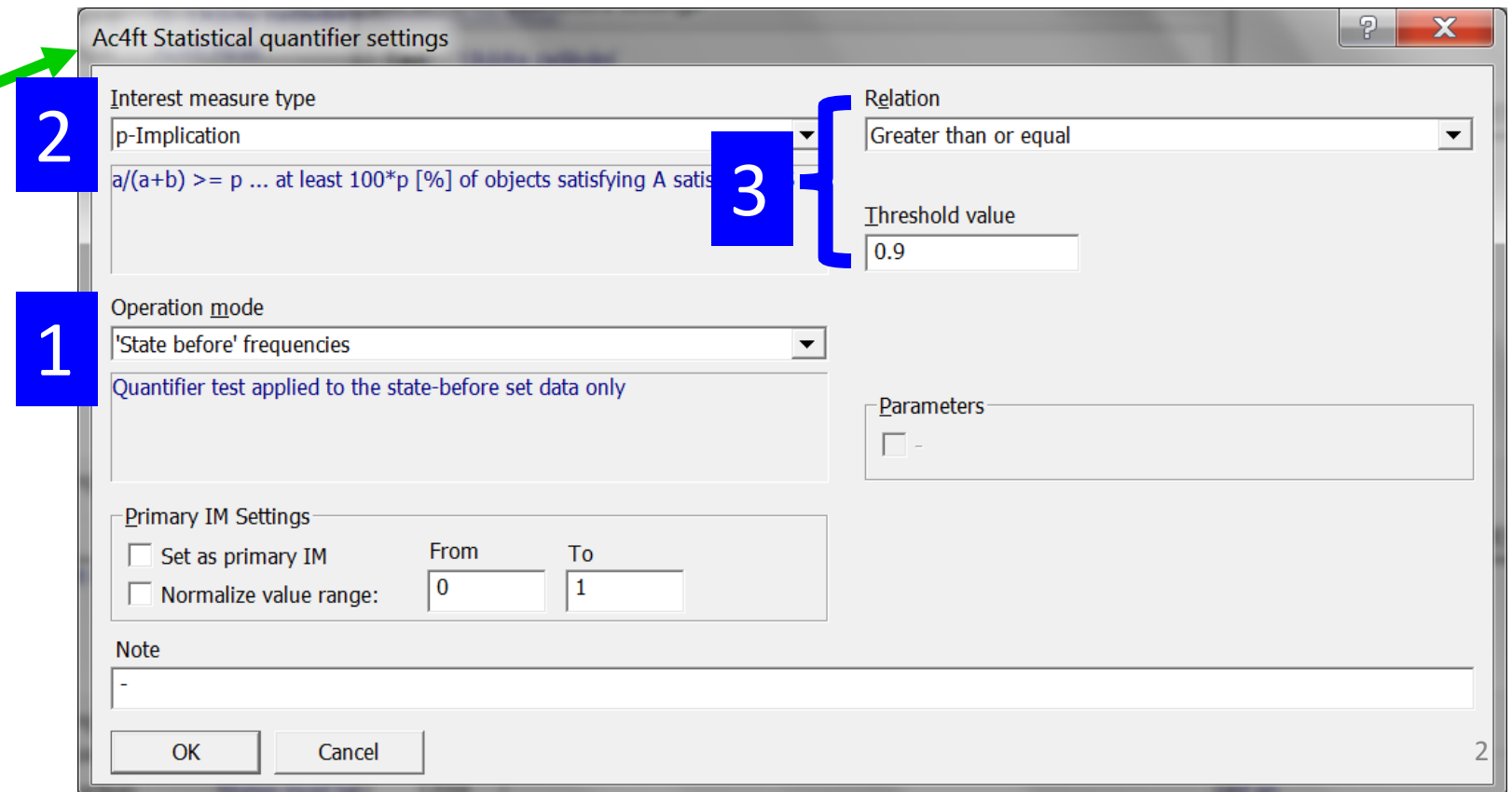
Katedra informačního a znalostního inženýrství

Vysoká škola ekonomická v Praze

Start tlačítkem Add simple frequencies quantifier



- Definice se provádí zadáním parametrů 1 - 3.
- Pro všechny parametry se nabízejí defaultní hodnoty.
- Podmínka kvantifikátoru je definována parametrem 3.



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

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

$$T_{Init} =$$

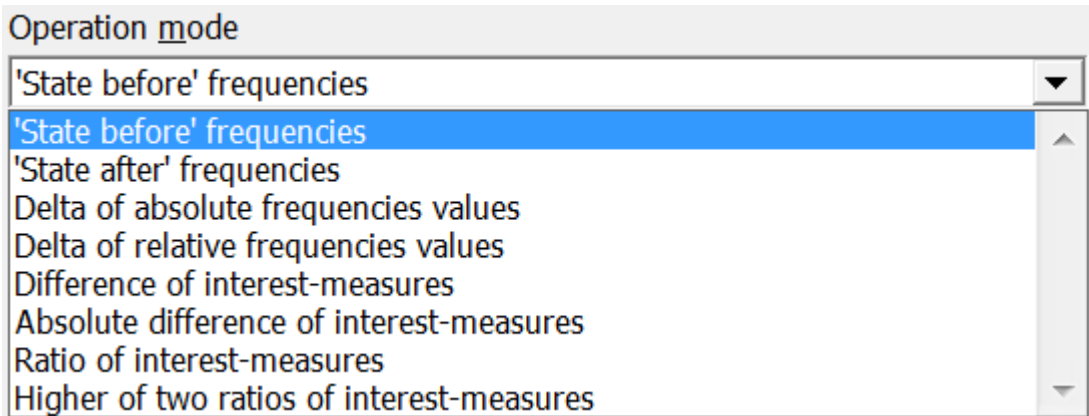
M/ χ	Ψ_{Init}	$\neg\Psi_{Init}$
ϕ_{Init}	a_I	b_I
$\neg\phi_{Init}$	c_I	d_I

$$T_{Fin} =$$

M/ χ	Ψ_{Fin}	$\neg\Psi_{Fin}$
ϕ_{Fin}	a_F	b_F
$\neg\phi_{Fin}$	c_F	d_F

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

$$\langle a, b, c, d \rangle = \langle \left| \frac{a_I}{n_I} - \frac{a_F}{n_F} \right|, \left| \frac{b_I}{n_I} - \frac{b_F}{n_F} \right|, \left| \frac{c_I}{n_I} - \frac{c_F}{n_F} \right|, \left| \frac{d_I}{n_I} - \frac{d_F}{n_F} \right| \rangle$$

$$\text{kde } n_I = a_I + b_I + c_I + d_I \text{ a } n_F = a_F + b_F + c_F + d_F.$$

2 - Interest measure type (A)

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

Interest measure type

p-Implication

p-Implication

Average Difference Dependence

Above Average Dependence

Below Average Dependence

Outside Average Dependence

Double p-Implication

p-Equivalence

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 parametrem Operation mode.

Interest measure type	Hodnota <i>IM</i>
<i>p</i> -Implication	$\frac{a}{a+b}$
Above Average Dependence	$\frac{a(a+b+c+d)}{(a+b)(a+c)} - 1$
Below Average Dependence	$1 - \frac{a(a+b+c+d)}{(a+b)(a+c)}$
Outside Average Dependence	$ \frac{a(a+b+c+d)}{(a+b)(a+c)} - 1 $
Average Difference Dependence	$\frac{\frac{a}{a+b}}{\frac{a}{a+c}} - 1$
Double <i>p</i> -implication	$\frac{a}{a+b+c}$
<i>p</i> -Equivalence	$\frac{a+d}{a+b+c+d}$

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*:

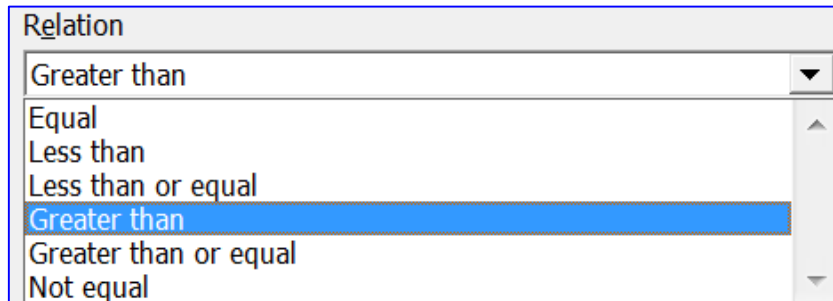
Interest measure type	IM_I	IM_F
p -Implication	$\frac{a_I}{a_I+b_I}$	$\frac{a_F}{a_F+b_F}$
Above Average Dependence	$\frac{a_I(a_I+b_I+c_I+d_I)}{(a_I+b_I)(a_I+c_I)} - 1$	$\frac{a_F(a_F+b_F+c_F+d_F)}{(a_F+b_F)(a_F+c_F)} - 1$
Below Average Dependence	$1 - \frac{a_I(a_I+b_I+c_I+d_I)}{(a_I+ha)(a_I+c_I)}$	$1 - \frac{a_F(a_F+b_F+c_F+d_F)}{(a_F+b_F)(a_F+c_F)}$
Outside Average Dependence	$ \frac{a_I(a_I+b_I+c_I+d_I)}{(a_I+b_I)(a_I+c_I)} - 1 $	$ \frac{a_F(a_F+b_F+c_F+d_F)}{(a_F+b_F)(a_F+c_F)} - 1 $
Average Difference Dependence	$\frac{\frac{a_I}{a_I+b_I}}{\frac{a_I+c_I}{(a_I+b_I+c_I+d_I)}} - 1$	$\frac{\frac{a_F}{a_F+b_F}}{\frac{a_F+c_F}{(a_F+b_F+c_F+d_F)}} - 1$
Double p -implication	$\frac{a_I}{a_I+b_I+c_I}$	$\frac{a_F}{a_F+b_F+c_F}$
p -Equivalence	$\frac{a_I+d_I}{a_I+b_I+c_I+d_I}$	$\frac{a_F+d_F}{a_F+b_F+c_F+d_F}$

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 - Relation x Threshold value

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



Relation

Greater than

Equal

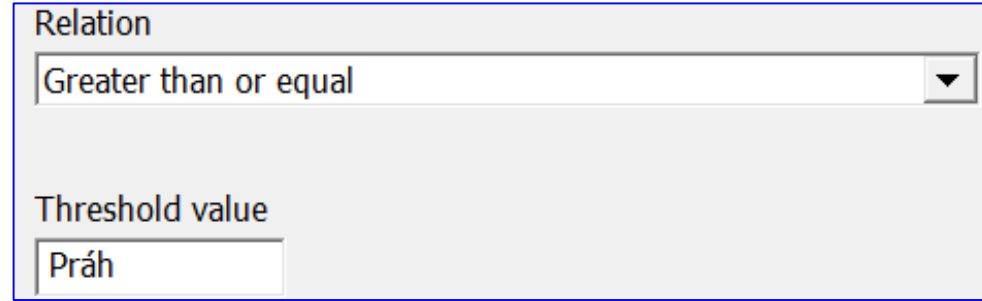
Less than

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 *IM* a hodnotou zadanou v poli *Threshold value* je považována za podmínku definující Ac4ft-kvantifikátor zadaný v krocích 1 až 3.