

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

Je dostupná z [této adresy](#)

Verze 26. 2. 2020

Typ úlohy: Aplikace jednoduchých frekvenčních
KL-kvantifikátorů, využití range u KL-kvantifikátoru

Data: [Hotel](#)

Problém: *Existuje kombinace hodnot atributů z Pobyť, Host, Bydliště, Meteo tak, že pro tuto kombinaci je počet hostů přijíždějících v pondělí až středa a hodnotících pobyt jako „průměr“ nebo „spokojen“ alespoň 50 %*

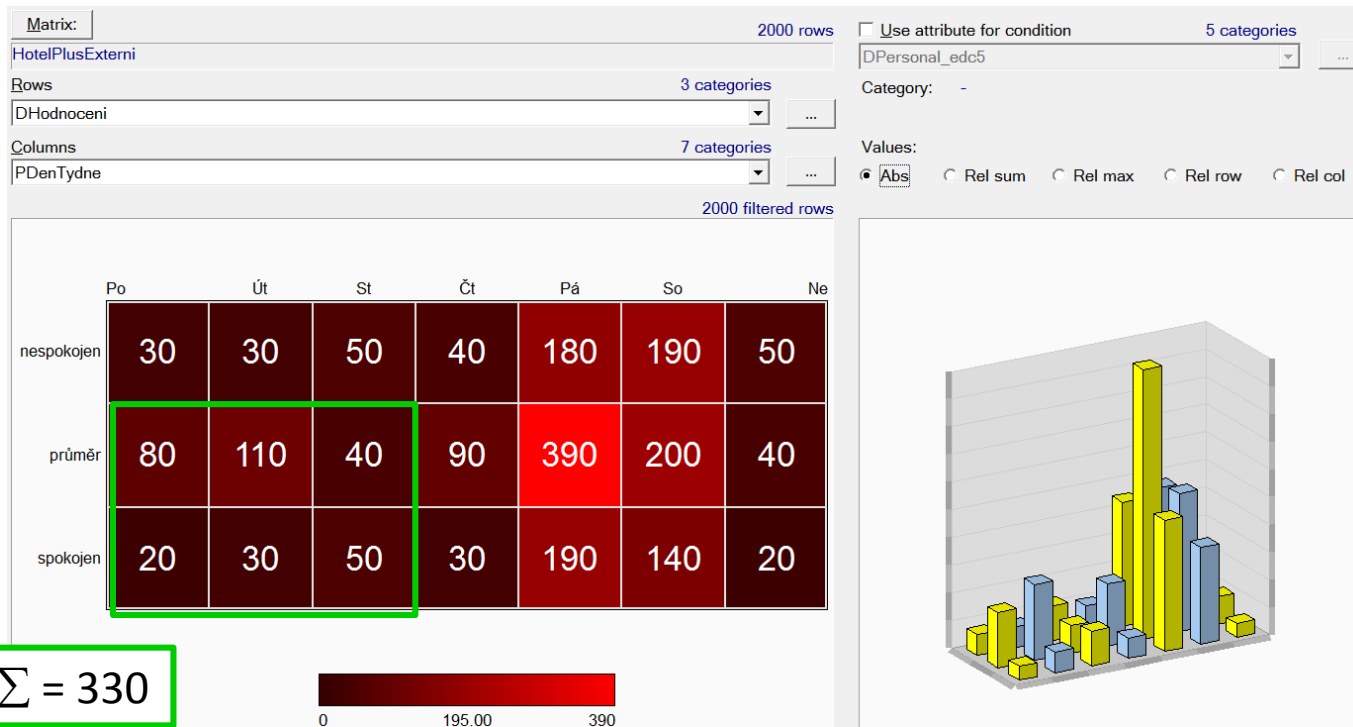
Jan Rauch

Katedra informačního a znalostního inženýrství

Vysoká škola ekonomická v Praze

Aplikace procedury KL-Miner – příklad

Inspirace: Počet hostů přijíždějících v pondělí až středa a hodnotících pobyt jako „průměr“ nebo „spokojen“ je celkem 330, t.j. 16.5 %.



Existuje kombinace hodnot atributů z Pobyt, Host, Bydliště, Meteo tak, že pro tuto kombinaci je počet hostů přijíždějících v pondělí až středa a hodnotících pobyt jako „průměr“ nebo „spokojen“ alespoň 50 % ?

Aplikace procedury KL-Miner – příklad 2

HVek	HPohlavi	HMesto	HMesto_X	HMesto_Y	HStat	PPobytOd	PNoci	POsob	PTypPobytu	PCenaUbytovani	PCenaStrava	PCenaSleva	PCenaCelkem	DHodnoceni
21	žena	České Budějovice	14.4757883	48.9763169	ČR	31.5.2013	1	1	rekreační	1450.00	0.000	0.00	1450.00	spokojen 91 82 81 56
34	muž	Linec	14.2862742	48.3066489	Rakousko	2.8.2013	2	4	rekreační	11600.00	1440.000	200.00	12840.00	průměr 44 21 62 84
30	muž	Vídeň	16.3736767	48.2115631	Rakousko	5.6.2012	7	2	rekreační	16940.00	2100.000	200.00	18840.00	nespokojen 5 37 25 71
62			1283											
35			67											
58			14											
81	žena	Vídeň	16.3736767											
22	žena	Dráždany	13.73970											
82	muž	Katovice	19.0241283											
55	muž	Praha	14.4212806	50										
75	žena	Berlín	13.3908886	52.5178189	Německo	26.3.2012	1	4	rekreační	4840.00	0.000	200.00	4840.00	nespokojen 30 39 32 3
66	žena	Linec	14.2862742	48.3066489	Rakousko	23.2.2012	1	2	služební	2420.00	0.000	0.00	2420.00	spokojen 95 93 75 21
64	žena	Linec	14.2862742	48.3066489	Rakousko	16.3.2013	7	2	rekreační	20300.00	0.000	600.00	19700.00	průměr 52 30 59 52
35	muž	Košice	21.2543528	48.7160408	Slovensko	6.4.2012	1	2	rekreační	2420.00	0.000	0.00	2420.00	průměr 36 52 56 36
32	muž	Mnichov	11.5836375	48.1364668	Německo	13.8.2013	1	1	služební	1450.00	180.000	0.00	1630.00	průměr 47 58 70 46
65														
79														
28														
35	žena	Hamburg	10.0043528	53.5498325	Německo	5.1.2013	14	2	rekreační	40600.00	5040.000	200.00	45440.00	nespokojen 27 15 30 12
22	muž	Plzeň	13.3771556	49.7490406	ČR	9.11.2013	4	1	rekreační	5800.00	0.000	200.00	5600.00	spokojen 84 79 86 48
25	muž	Karlovy Vary	12.8690381	50.2311075	ČR	9.11.2013	7	2	rekreační	20300.00	0.000	600.00	19700.00	spokojen 88 74 94 54
20	žena	Hamburg	10.0043528	53.5498325	Německo	19.1.2013	14	4	rekreační	81200.00	10080.000	200.00	91080.00	průměr 59 23 46 96
30	žena	Linec				29.11.2013	2	4	rekreační	11600.00	0.000	200.00	11400.00	spokojen 88 82 84 61
45	žena	Karlovy				28.12.2012	2	2	rekreační	4840.00	600.000	200.00	5240.00	spokojen 81 99 87 23

Hotel.txt

Existuje kombinace hodnot atributů z Pobyt, Host, Bydliště, Meteo tak, že pro tuto kombinaci je počet hostů přijíždějících v pondělí až středa a hodnotících pobyt jako „průměr“ nebo „spokojen“ alespoň 50 % ?

? KL: SUM_{2:1..3:3, ≥50%} DHodnoceni x PDenTydne / Pobyt, Host, Bydliště, Meteo

Meteo.txt

MDatum	MTeplota	MOblaha
4.1.2012	-6.3	slunečno
5.1.2012	-6.6	zataženo
6.1.2012	6.1	srážky
7.1.2012	1.6	srážky
8.1.2012	-1.3	srážky
		7 zataženo
		srážky
		zataženo
12.1.2012	-3.1	srážky
13.1.2012	-8.1	zataženo
14.1.2012	-10.7	srážky
15.1.2012	-5.5	zataženo
16.1.2012	2.3	zataženo
17.1.2012	-1.9	zataženo
18.1.2012	-8.6	zataženo

? KL: $SUM_{2:1..3:3, 50\%}$ Phodnoceni x PDenTydne / Pobyt, Host, Bydliště, Meteo

ROW ATTRIBUTES	QUANTIFIERS	COLUMN ATTRIBUTES																														
Row attributes » DHodnoceni(3 categories)	<table border="1"> <thead> <tr> <th>Type</th> <th>Range</th> <th>Rel.</th> <th>Value</th> <th>Units</th> </tr> </thead> <tbody> <tr> <td>SUM</td> <td>all</td> <td>>=</td> <td>100.00</td> <td>Abs</td> </tr> <tr> <td>SUM</td> <td>2:1..3:3</td> <td>>=</td> <td>50.00</td> <td>%Cond</td> </tr> </tbody> </table>	Type	Range	Rel.	Value	Units	SUM	all	>=	100.00	Abs	SUM	2:1..3:3	>=	50.00	%Cond	Column attributes » PDenTydne(7 categories)															
Type	Range	Rel.	Value	Units																												
SUM	all	>=	100.00	Abs																												
SUM	2:1..3:3	>=	50.00	%Cond																												
DHodnocení	$SUM \geq 100 \wedge SUM_{2:1..3:3, \geq 50\%}$	PDenTydne																														
Aggregate function Type: Count(*) Attribute: -	Generation information Status: Solved, 25 run(s) Mode: Standard	<table border="1"> <thead> <tr> <th colspan="2">CONDITION</th> </tr> </thead> <tbody> <tr> <td>Pobyt</td> <td>Con, 1 - 4</td> </tr> <tr> <td>» PNoci_enum_m (seq), 1 - 2</td> <td>B, pos</td> </tr> <tr> <td>» POsob (seq), 1 - 3</td> <td>B, pos</td> </tr> <tr> <td>» POSobonoci_ef5 (seq), 1 - 2</td> <td>B, pos</td> </tr> <tr> <td>Host</td> <td>Con, 0 - 2</td> </tr> <tr> <td>» HPohlavi (subset), 1 - 1</td> <td>B, pos</td> </tr> <tr> <td>» HVek_exp (seq), 1 - 2</td> <td>B, pos</td> </tr> <tr> <td>Bydliště</td> <td>Con, 0 - 1</td> </tr> <tr> <td>» HCizinec_b (subset), 1 - 1</td> <td>B, pos</td> </tr> <tr> <td>» HMesto (subset), 1 - 1</td> <td>B, pos</td> </tr> <tr> <td>» HStat (subset), 1 - 1</td> <td>B, pos</td> </tr> <tr> <td>Meteo</td> <td>Con, 0 - 5</td> </tr> <tr> <td>» MObloha (subset), 1 - 1</td> <td>B, pos</td> </tr> <tr> <td>» MTEplota_exp (seq), 1 - 2</td> <td>B, pos</td> </tr> </tbody> </table>	CONDITION		Pobyt	Con, 1 - 4	» PNoci_enum_m (seq), 1 - 2	B, pos	» POsob (seq), 1 - 3	B, pos	» POSobonoci_ef5 (seq), 1 - 2	B, pos	Host	Con, 0 - 2	» HPohlavi (subset), 1 - 1	B, pos	» HVek_exp (seq), 1 - 2	B, pos	Bydliště	Con, 0 - 1	» HCizinec_b (subset), 1 - 1	B, pos	» HMesto (subset), 1 - 1	B, pos	» HStat (subset), 1 - 1	B, pos	Meteo	Con, 0 - 5	» MObloha (subset), 1 - 1	B, pos	» MTEplota_exp (seq), 1 - 2	B, pos
CONDITION																																
Pobyt	Con, 1 - 4																															
» PNoci_enum_m (seq), 1 - 2	B, pos																															
» POsob (seq), 1 - 3	B, pos																															
» POSobonoci_ef5 (seq), 1 - 2	B, pos																															
Host	Con, 0 - 2																															
» HPohlavi (subset), 1 - 1	B, pos																															
» HVek_exp (seq), 1 - 2	B, pos																															
Bydliště	Con, 0 - 1																															
» HCizinec_b (subset), 1 - 1	B, pos																															
» HMesto (subset), 1 - 1	B, pos																															
» HStat (subset), 1 - 1	B, pos																															
Meteo	Con, 0 - 5																															
» MObloha (subset), 1 - 1	B, pos																															
» MTEplota_exp (seq), 1 - 2	B, pos																															
Task parameters Include 'worse' extensions of condition: <input checked="" type="checkbox"/> Yes Extensions minimal length check: <input checked="" type="checkbox"/> Yes Include extensions of coefficients with no change in the histogram: <input checked="" type="checkbox"/> Yes Include extensions of cedents with no change in the histogram: <input checked="" type="checkbox"/> Yes Maximal number:	Jedno z více možných zadání pro řešení dané analytické otázky	$Pobyt (*) \wedge Host (*) \wedge$ $\wedge Bydliště(*) \wedge Meteo(*)$																														

KL-kvantifikátor SUM all

QUANTIFIERS				
Type	Range	Rel.	Value	Units
SUM	all	>=	100.00	Abs
SUM	2:1..3:3	>=	50.00	%Cond

KL Simple frequencies quantifier settings

Interest measure type: **Sum of frequencies**
Sum of frequencies from given part of contingency table

Relation: Greater than or equal

Threshold value: 100

Threshold-value units: Absolute number
Absolute number of rows. Threshold value is left as given.

Category Range

	From	To	
Rows	0	100	Reset to All
Columns	0	100	

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

Primary IM Settings

Set as primary IM

	From	To
<input type="checkbox"/> Normalize value range:	0	1

Note: -

OK Cancel

SUM >= 100 minimálně 100 řádků splňuje podmínku

KL-kvantifikátor SUM 2:1..3:3

QUANTIFIERS				
Type	Range	Rel.	Value	Units
SUM	all	>=	100.00	Abs
SUM	2:1..3:3	>=	50.00	%Cond

KL Simple frequencies quantifier settings

Interest measure type: **Sum of frequencies**
 Sum of frequencies from given part of contingency table

Relation: Greater than or equal

Threshold value: 50

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

Category Range

	From	To
Rows	2	3
Columns	1	3

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

OK Cancel

	Po	Út	St	Ct	Pá	So	Ne
nepokojen	30	30	50	40	180	190	50
průměr	80	110	40	90	390	200	40
spokojen	20	30	50	30	190	140	20

KL-Miner – příklad výstupu

Task run

Start: 28.2.2016 20:24:05 Total time: 0h 0m 2s

Number of verifications: 18950 Mode: Standard

Number of hypotheses: 88

Add group Del group Edit group

Actual group of hypotheses: All hypotheses

Hypotheses in group: 88 Shown hypotheses: 88 Highlighted: 0

Nr.	Id	Sum	Hypothesis
1	29	146	DHodnoceni × PDenTydne / PNoci(1) & POsob(2,3) & POsobonoci(nižší) & HCizinec(ne)
2	30	146	DHodnoceni × PDenTydne / PNoci(1) & POsob(2,3) & POsobonoci(nižší) & HStat(ČR)
3	25	146	DHodnoceni × PDenTydne / PNoci(1) & POsob(2,3) & POsobonoci(<=nižší) & HCizinec(ne)
4	26	146	DHodnoceni × PDenTydne / PNoci(1) & POsob(2,3) & POsobonoci(<=nižší) & HStat(ČR)
5	33	146	DHodnoceni × PDenTydne / PNoci(1) & POsob(2,3) & POsobonoci(nižší,průměr) & HCizinec(ne)
6	34	146	DHodnoceni × PDenTydne / PNoci(1) & POsob(2,3) & POsobonoci(nižší,průměr) & HStat(ČR)
7	21	146	DHodnoceni × PDenTydne / PNoci(1) & POsob(2,3) & HCizinec(ne)
8	22	146	DHodnoceni × PDenTydne / PNoci(1) & POsob(2,3) & HStat(ČR)
9	5	146	DHodnoceni × PDenTydne / PNoci(1) & POsob(<=3) & POsobonoci(nižší) & HCizinec(ne)
10	6	146	DHodnoceni × PDenTydne / PNoci(1) & POsob(<=3) & POsobonoci(nižší) & HStat(ČR)
11	9	146	DHodnoceni × PDenTydne / PNoci(1) & POsob(<=3) & POsobonoci(nižší,průměr) & HCizinec(ne)
12	10	146	DHodnoceni × PDenTydne / PNoci(1) & POsob(<=3) & POsobonoci(nižší,průměr) & HStat(ČR)
13	41	146	DHodnoceni × PDenTydne / PNoci(1) & POsob(>=2) & POsobonoci(nižší) & HCizinec(ne)
14	42	146	DHodnoceni × PDenTydne / PNoci(1) & POsob(>=2) & POsobonoci(nižší) & HStat(ČR)
15	37	146	DHodnoceni × PDenTydne / PNoci(1) & POsob(>=2) & POsobonoci(<=nižší) & HCizinec(ne)
16	38	146	DHodnoceni × PDenTydne / PNoci(1) & POsob(>=2) & POsobonoci(<=nižší) & HStat(ČR)
17	45	146	DHodnoceni × PDenTydne / PNoci(1) & POsobonoci(nižší) & HCizinec(ne)
18	46	146	DHodnoceni × PDenTydne / PNoci(1) & POsobonoci(nižší) & HStat(ČR)
19	57	146	DHodnoceni × PDenTydne / PNoci(<=2) & POsob(2,3) & POsobonoci(nižší) & HCizinec(ne)

KL-Miner – příklad detailního výstupu

DHodnoceni x PDenTydne / PNoci(1) \wedge POsob(2,3) \wedge POsobonoci(nižší) \wedge HCizinec(ne)
SUM = 146

Attributes: DHodnoceni x PDenTydne: frequencies
Condition: PNoci(1) & POsob(2, 3) & POsobonoci(nižší) & HCizinec(ne)

TEXT | DATA | FREQUENCIES

	Po	Út	St	Čt	Pá	So	Ne
nespokojen	1	2	6	5	0	1	5
průměr	27	20	3	16	8	11	3
spokojen	0	15	10	1	9	2	1

$$\Sigma = 75$$

$$75/146 = 51,3 \%$$

