

Interaktivní analýza dat v systému LIsp-Miner

prof. RNDr. Jan Rauch, CSc.

Katedra informačního a znalostního inženýrství

Interaktivní analýza dat v systému LIsp-Miner

- Úvod
- Příklady:
 - histogram na celých datech
 - histogram na podmnožině dat
 - kontingenční tabulka na celých datech
 - kontingenční tabulka na podmnožině dat
 - asociační pravidlo

Interaktivní analýza dat - úvod

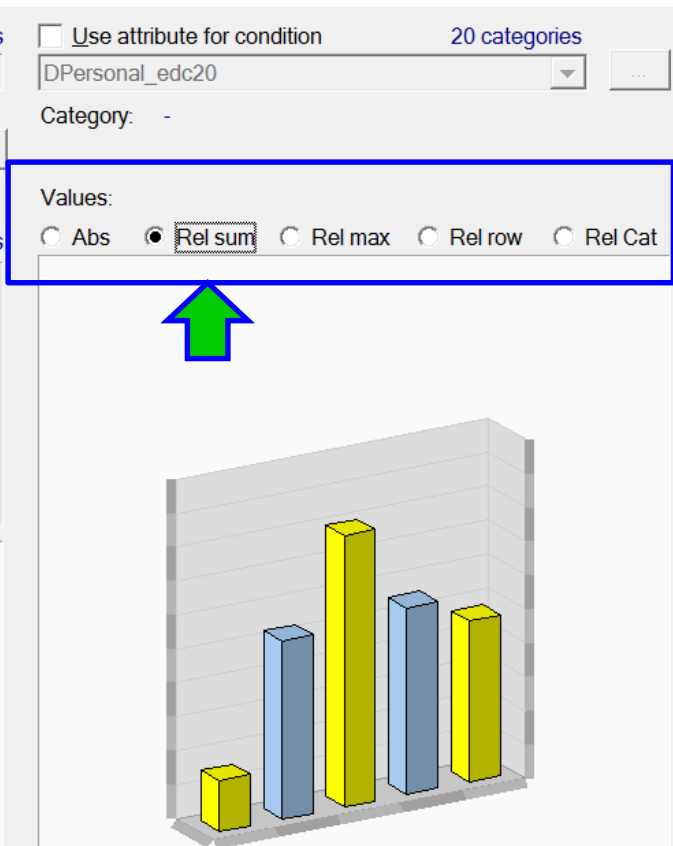
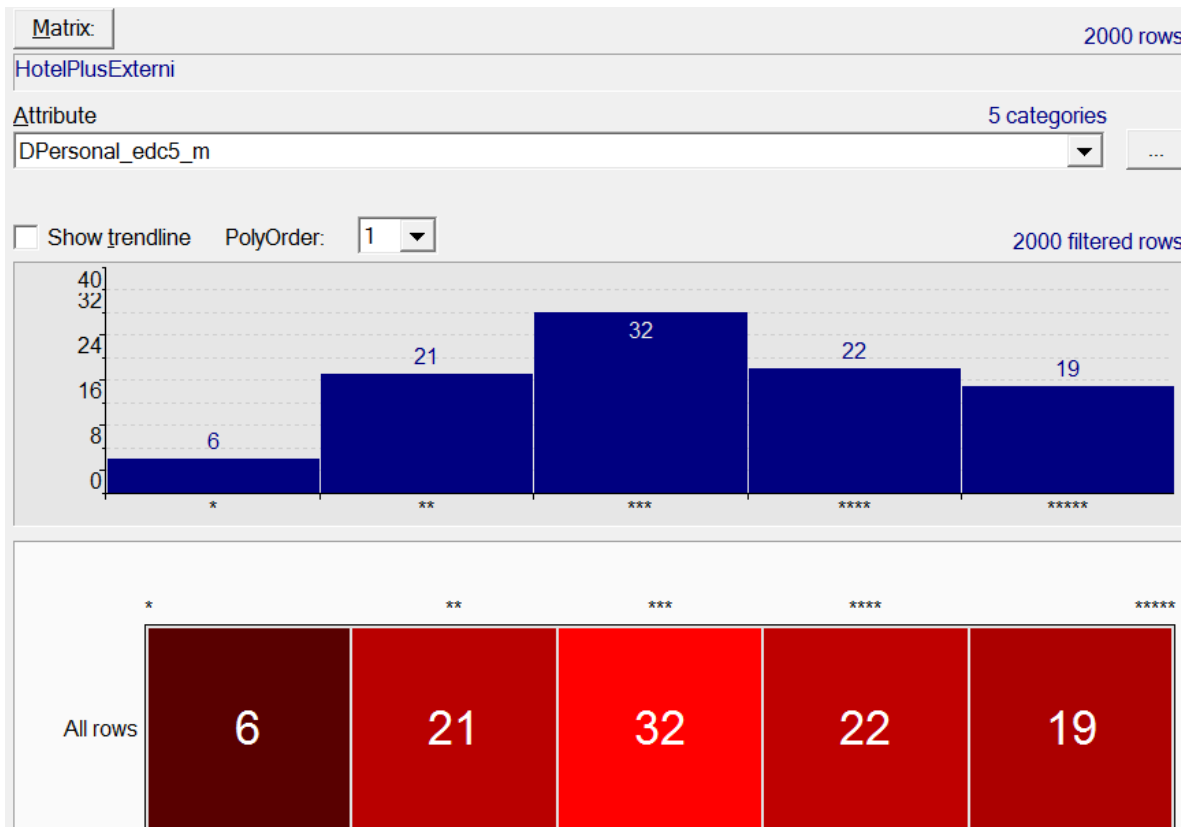
Interactive Analysis	Data-mining Tasks	Domain Knowledge
New Scatter-plot (XY) Analysis Tab		
New Principal Component Analysis (PCA) Tab		
New Category Frequency Analysis Tab		
New KL Attributes Correlation Analysis Tab		
New CF Contingency Analysis Tab		
New KL Contingency Analysis Tab		
New 4ft Contingency Analysis Tab		
New ET Decision Tree Analysis (TDIDT) Tab		
New Geodata XY Analysis Tab		
New Geodata CF Analysis Tab		
Gallery of Stored Patterns Tab		

- Užitečný nástroj pro seznámení s daty a pro hledání důležitých silných vztahů
- Podrobný popis na <https://lispminer.vse.cz/wiki/doku.php?id=lmianalysis:start>
- Dále uvedeny příklady využití pro výpočet globálních a některých silných vztahů (globálním vztahem rozumíme vztah na celých datech a silným vztahem rozumíme vztah na zajímavé a dostatečně rozsáhlé podmnožině dat).
- Pro výpočet silných vztahů je však možno využít i GUHA - procedury.

Histogram na celých datech

Hodnocení personálu pro všech 2 000 pobytů

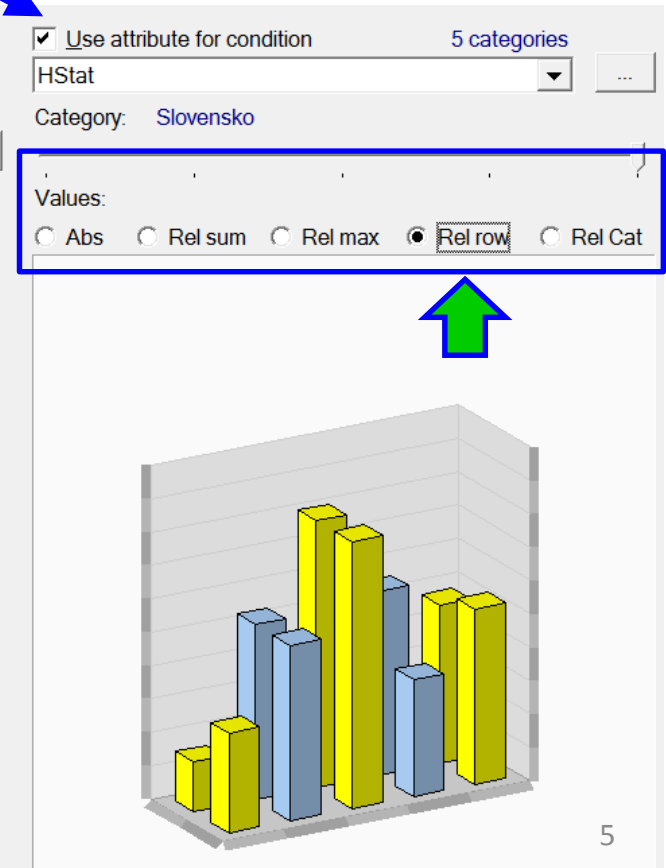
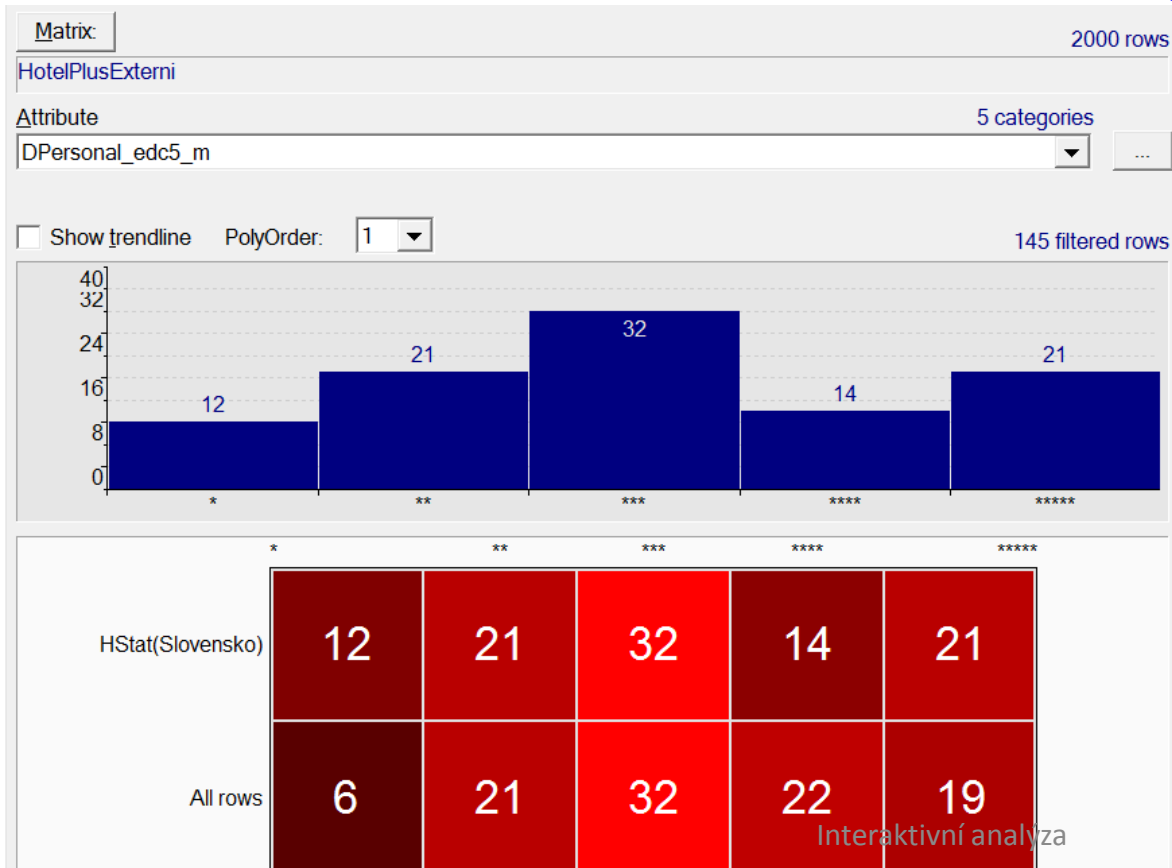
- Interactive Analysis
- Data-mining Tasks
- Domain Knowl
- New Scatter-plot (XY) Analysis Tab
- New Principal Component Analysis (PCA) Tab
- New Category Frequency Analysis Tab
- New KL Attributes Correlation Analysis Tab
- New CF Contingency Analysis Tab
- New KL Contingency Analysis Tab
- New 4ft Contingency Analysis Tab
- New ET Decision Tree Analysis (TDIDT) Tab
- New Geodata XY Analysis Tab
- New Geodata CF Analysis Tab
- Gallery of Stored Patterns Tab



Histogram na podmnožině dat

Hodnocení personálu pro 145 pobytů ze Slovenska

- Interactive Analysis
- Data-mining Tasks
- Domain Knowl
- New Scatter-plot (XY) Analysis Tab
- New Principal Component Analysis (PCA) Tab
- New Category Frequency Analysis Tab
- New KL Attributes Correlation Analysis Tab
- New CF Contingency Analysis Tab
- New KL Contingency Analysis Tab
- New 4ft Contingency Analysis Tab
- New ET Decision Tree Analysis (TDIDT) Tab
- New Geodata XY Analysis Tab
- New Geodata CF Analysis Tab
- Gallery of Stored Patterns Tab



Kontingenční tabulka na celých datech

Vztah **Hodnocení personálu X Hodnocení ubytování**
pro všech 2 000 pobytů



- Interactive Analysis
- Data-mining Tasks
- Domain Knowl
- New Scatter-plot (XY) Analysis Tab
- New Principal Component Analysis (PCA) Tab
- New Category Frequency Analysis Tab
- New KL Attributes Correlation Analysis Tab
- New CF Contingency Analysis Tab
- New KL Contingency Analysis Tab
- New 4ft Contingency Analysis Tab
- New ET Decision Tree Analysis (TDIDT) Tab
- New Geodata XY Analysis Tab
- New Geodata CF Analysis Tab
- Gallery of Stored Patterns Tab

Matrix: 2000 rows

HotelPlusExterni

Rows: 5 categories

DPersonal_edc5_m

Columns: 5 categories

DUbytovani_edc5_m

2000 filtered rows

Use attribute for condition 5 categories

DPersonal_edc5

Category: -

Values:

Abs Rel sum Rel max Rel row Rel col

	*	**	***	****	*****
*	28	88	4	0	0
**	131	270	18	0	0
***	13	147	453	30	0
****	0	40	274	24	95
*****	0	0	0	0	385

Interaktivní analýza

6

Kontingenční tabulka na podmnožině dat

Vztah **Hodnocení personálu X Hodnocení ubytování**
pro 480 pobytů s hodnocením DHodnocení = spokojen

- Interactive Analysis
- Data-mining Tasks
- Domain Knowl
- New Scatter-plot (XY) Analysis Tab
- New Principal Component Analysis (PCA) Tab
- New Category Frequency Analysis Tab
- New KL Attributes Correlation Analysis Tab
- New CF Contingency Analysis Tab
- New KL Contingency Analysis Tab
- New 4ft Contingency Analysis Tab
- New ET Decision Tree Analysis (TDIDT) Tab
- New Geodata XY Analysis Tab
- New Geodata CF Analysis Tab
- Gallery of Stored Patterns Tab

Matrix: 2000 rows

HotelPlusExterni

Rows: 5 categories
DPersonal_edc5_m

Columns: 5 categories
DUbytovani_edc5_m

480 filtered rows

Use attribute for condition 3 categories
DHodnoceni

Category: spokojen

Values:
 Abs Rel sum Rel max Rel row Rel col

	*	**	***	****	*****
*	0	0	0	0	0
**	0	0	0	0	0
***	0	0	0	0	0
****	0	0	0	0	20
*****	0	0	0	0	80

Asociační pravidlo

- Interactive Analysis
- Data-mining Tasks
- Domain Knowledge
- New Scatter-plot (XY) Analysis Tab
- New Principal Component Analysis (PCA) Tab
- New Category Frequency Analysis Tab
- New KL Attributes Correlation Analysis Tab
- New CF Contingency Analysis Tab
- New KL Contingency Analysis Tab
- New 4ft Contingency Analysis Tab
- New ET Decision Tree Analysis (TDIDT) Tab
- New Geodata XY Analysis Tab
- New Geodata CF Analysis Tab
- Gallery of Stored Patterns Tab

Ukázka inspirovaná pravým sloupcem v předchozím příkladu.
Odpovídá pravidlu

DHodnocení(spokojen) \wedge Hodnocení ubytování(****) \Rightarrow 0,8,385 Hodnocení personálu(****)

Matrix: 2000 rows

HotelPlusExterni

Rows: 5 categories
DPersonal_edc5_m

Columns: 5 categories
DUbytovani_edc5_m

480 filtered rows

	*	**	***	****	*****
*	0	0	0	0	0
**	0	0	0	0	0
***	0	0	0	0	0
****	0	0	0	0	20
*****	0	0	0	0	80

Use attribute for condition: 3 categories
DHodnoceni

Category: spokojen

Values:
 Abs Rel sum Rel max Rel row Rel col

Asociační pravidlo - konstrukce antecedentu

Matrix: HotelPlusExterni

Antecedent: (empty)

TEXT | DATA | GRAPH | 9-FOLD TABLE

Succedent

Antecedent	2000	
¬Antecedent	0	

0 1000.00

Value: Abs Rel sum Rel max Rel row Rel col

Antecedent Succedent Condition Store to Gallery Export

Tree of attributes

Data matrix: HotelPlusExterni

All attributes in the selected matrix

- DHodnoceni
 - nespokojen
 - průměr
- DPersonal_edc20
- DPersonal_edc5
- DPersonal_edc5_m
- DPersonal_ef3
- DStrava_edc20
- DStrava_edc5
- DStrava_edc5_m
- DStrava_ef3
- DUbytovani_edc20
- DUbytovani_edc5
- DUbytovani_edc5_m**
- *
- **
- ***
- ****
- DUbytovani_ef3
- DZabava_edc20

Defined cedent

- Conjunction of literals
 - DHodnoceni
 - spokojen
 - DUbytovani
 - *****

OK Cancel Dgfail Expand all Empty



Asociační pravidlo - konstrukce sukcedentu

The screenshot displays a software interface for analyzing association rules. The main window shows a matrix with the following data:

	Succedent
Antecedent	480
\neg Antecedent	1520

The matrix also shows a color scale at the bottom ranging from 0 to 760.00. The 'Tree of attributes' dialog is open, showing a hierarchy of attributes. The selected attribute is 'DPersonal_edc5_m'. The dialog also shows a 'Defined cedent' section with 'DPersonal' and '*****'.

Value: Abs Rel sum Rel max Rel row Rel col

Antecedent **Succedent** Condition Store to Gallery Export



Asociační pravidlo - výsledek

DHodnocení(spokojen) \wedge Hodnocení ubytování(*****) $\Rightarrow_{0.8,385}$ Hodnocení personálu(*****)

$$385 / (385 + 95) = 0.8$$

