Tato prezentace je součástí wiki-prezentace Metoda GUHA, LISp-Miner a typové úlohy
Je dostupná z této adresy
Verse 17. 2. 2020

Typ úlohy: Asociační pravidla popisující výjimky (Exception rules)

Data: Accidents

Problém: Existuje podmínka P, dodatečná podmínka D a závažnost Z nehody takové, že

- relativní četnost nehod se závažností Z za podmínky P je vyšší než relativní četnost poruch se závažnosti Z mezi všemi nehodami
- relativní četnost nehod se závažností Z je za současné platnosti podmínky P i dodatečné podmínky D nižší, než je relativní četnost nehod se závažnosti Z mezi všemi nehodami ?

Jan Rauch
Katedra informačního a znalostního inženýrství
Vysoká škola ekonomická v Praze

© Jan Rauch

Nadprůměrná a podprůměrná relativní četnost

- Příklad
- Podmínky, za kterých je relativní četnost některé závažnosti nadprůměrná
- Dodatečné podmínky, za kterých je relativní četnost stejné závažnosti podprůměrná
- Princip řešení
- Poznámky

Příklad

Podmínka **P**: Sex(Male) ∧ Engine_Capacity__CC_⟨4,1199)

Závažnost **Z**: Accident_Severity(Serious)

Dodatečná podmínka **D**: Speed_limit(>=50) ∧ Vehicle_type(Car)

	Z	¬Z
P	16016	40520
¬P	89738	392715

	Z	¬Z
P∧D	1278	6611
¬ (P∧D)	104476	426624

$$\frac{16016}{16016+40520} = (1+0.44)\frac{16016+89738}{16016+40520+89738+392715}$$

$$\frac{1278}{1278+6611} = (1-0.17)\frac{1278+104476}{1278+6611+104476+426624}$$

- relativní četnost nehod se závažností **Z** za podmínky **P** je o **44%** vyšší než relativní četnost poruch se závažnosti **Z** mezi všemi nehodami
- relativní četnost nehod se závažností **Z** je za současné platnosti **P** podmínky i dodatečné podmínky **D** je o **17%** nižší, než je relativní četnost nehod se závažnosti **Z** mezi všemi nehodami

Podmínky za kterých je relativní četnost některé závažnosti nadprůměrná

- Zadání parametrů pro proceduru 4ft-Miner
- Přehled výsledků
- Nejsilnější pravidlo - detail

Zadání parametrů pro proceduru 4ft-Miner (1)

ANTECEDENT

Accident specifics	Con, 0 - 5
» Area_type (subset), 1 - 1	B, pos
» 1st_Road_Class (subset), 1 - 1	B, pos
» Carriageway_Hazards (subset), 1 - 1	B, pos
» Junction_Detail (subset), 1 - 1	B, pos
» Light_Conditions (subset), 1 - 1	B, pos
» Road_Surface_Conditions (subset), 1 - 1	B, pos
» Speed_limit _ed3 (subset), 1 - 1	B, pos
» Weather_Conditions (subset), 1 - 1	B, pos
Driver	Con, 0 - 5
» Age_Groups_of_Driver (subset), 1 - 1	B, pos
» Driver_Home_Area_Type (subset), 1 - 1	B, pos
» Driver_IMD_Decile (seq), 1 - 2	B, pos
» Journey_Purpose_of_Driver (subset), 1 - 1	B, pos
» Sex_of_Driver (subset), 1 - 1	B, pos

Total length: 1 - 2

QUANTIFIERS

BASE	p= 10000 Abs.
AAD	p= 0.200

Generation information

Status: Solved, 10 run(s)

Mode: Standard

SUCCEDENT

Default Partial Cedent	Con, 1 - 5
» Accident_Severity (subset), 1 - 1	B, pos

Total length: 1

- Minimálně 10 000 nehod
- Minimálně o 20% vyšší relativní četnost pro některou závažnost poruchy

Basic parameters

Name: Driver

Min. length: 0 Max. length: 5 Literals boolean operation type: Conjunction

Edit

Comment: -

Options

Allow only a consecutive sequence of literals in cedent (only neighbouring literals): Yes

Linked coefficients (all literals must have the same coefficient as in the first one): No

Literals Settings

Underlying attribute	Categories	X-cat	Coefficient type	Length	+/-	B/R	Class of equiv.
Age_Groups_of_Driver	9	Yes	Subsets	1 - 1	pos	Basic	-
Driver_Home_Area_Type	3	Yes	Subsets	1 - 1	pos	Basic	-
Driver_IMD_Decile	10	Yes	Sequences	1 - 2	pos	Basic	-
Journey_Purpose_of_Driver	5	Yes	Subsets	1 - 1	pos	Basic	-
Sex_of_Driver	2	Yes	Subsets	1 - 1	pos	Basic	-

Zadání parametrů pro proceduru 4ft-Miner (2)

Basic parameters

Name: Time

Min. length: 0 Max. length: 5 Literals boolean operation type: Conjunction Edit

Comment: -

Options

Allow only a consecutive sequence of literals in cedent (only neighbouring literals): No

Linked coefficients (all literals must have the same coefficient as in the first one): No

Literals Settings

Underlying attribute	Categories	X-cat	Coefficient type	Length	+/-	B/R	Class of equiv.
Day_of_Week	7	No	Subsets	1 - 1	pos	Basic	Day_of_week__Weekend
Weekend	2	No	Subsets	1 - 1	pos	Basic	Day_of_week__Weekend
Hour	24	Yes	Subsets	1 - 1	pos	Basic	Hours
Rush_hours	2	Yes	Subsets	1 - 1	pos	Basic	Hours
Month	12	No	Subsets	1 - 1	pos	Basic	Quarter_Month
Quarter	4	No	Subsets	1 - 1	pos	Basic	Quarter_Month
Year	11	No	Subsets	1 - 1	pos	Basic	-

Basic parameters

Name: Vehicle

Min. length: 0 Max. length: 5 Literals boolean operation type: Conjunction Edit

Comment: -

Options

Allow only a consecutive sequence of literals in cedent (only neighbouring literals): No

Linked coefficients (all literals must have the same coefficient as in the first one): No

Literals Settings

Underlying attribute	Categories	X-cat	Coefficient type	Length	+/-	B/R	Class of equiv.
Age_of_Vehicle_exp	5	Yes	Subsets	1 - 1	pos	Basic	-
Engine_Capacity__CC_	5	Yes	Subsets	1 - 1	pos	Basic	-
Vehicle_Type 5000 +	10	Yes	Subsets	1 - 1	pos	Basic	-
Was_Vehicle_Left_Hand_Drive	2	Yes	Subsets	1 - 1	pos	Basic	-
Propulsion_Code 5000+	2	Yes	Subsets	1 - 1	pos	Basic	-

Přehled výsledků

Task run

Start: 15.2.2020 00:19:04 Total time: 0h 2m 44s

Number of verifications: 12522

Number of hypotheses: 15 Mode: Standard

Add group | Del group | Edit group

Actual group of hypotheses: All hypotheses

Hypotheses in group: 15 Shown hypotheses: 15 Highlighted: 0

Nr.	Id	AAD	Hypothesis
1	6	0.444	**Sex**(*Male*) & **Engine_Capacity__CC_**(< *1199*) >÷< **Accident_Severity**(*Serious*)
2	12	0.438	**Road_Surface_Conditions**(*Dry*) & **Engine_Capacity__CC_**(< *1199*) >÷< **Accident_Severity**(*Serious*)
3	8	0.399	**Area**(*Rural*) & **Engine_Capacity__CC_**(< *1199*) >÷< **Accident_Severity**(*Serious*)
4	15	0.360	**Weather_Conditions**(*Fine no high winds*) & **Engine_Capacity__CC_**(< *1199*) >÷< **Accident_Severity**(*Serious*)
5	10	0.314	**Junction_Detail**(*Not at junction or within 20 metres*) & **Engine_Capacity__CC_**(< *1199*) >÷< **Accident_Severity**(*Serious*)
6	4	0.311	**Rush**(*Other*) & **Engine_Capacity__CC_**(< *1199*) >÷< **Accident_Severity**(*Serious*)
7	11	0.287	**Light_Conditions**(*Daylight*) & **Engine_Capacity__CC_**(< *1199*) >÷< **Accident_Severity**(*Serious*)
8	5	0.284	**Driver_Home_Area_Type**(*Urban area*) & **Engine_Capacity__CC_**(< *1199*) >÷< **Accident_Severity**(*Serious*)
9	2	0.279	**Engine_Capacity__CC_**(< *1199*) & **Propulsion_Code**(*Petrol*) >÷< **Accident_Severity**(*Serious*)
10	3	0.278	**Engine_Capacity__CC_**(< *1199*) & **Left_Hand_Drive**(*No*) >÷< **Accident_Severity**(*Serious*)
11	9	0.277	**Carriageway_Hazards**(*None*) & **Engine_Capacity__CC_**(< *1199*) >÷< **Accident_Severity**(*Serious*)
12	1	0.275	**Engine_Capacity__CC_**(< *1199*) >÷< **Accident_Severity**(*Serious*)
13	14	0.234	**Road_Surface_Conditions**(*Dry*) & **Speed_limit**(>= *50*) >÷< **Accident_Severity**(*Serious*)
14	13	0.234	**Road_Surface_Conditions**(*Dry*) & **Day**(*Sunday*) >÷< **Accident_Severity**(*Serious*)
15	7	0.219	**Sex**(*Male*) & **Day**(*Sunday*) >÷< **Accident_Severity**(*Serious*)

Nejsilnější pravidlo - detail

Podmínka **P**: Sex(Male) $\wedge$ Engine_Capacity__CC_$\langle$4,1199)

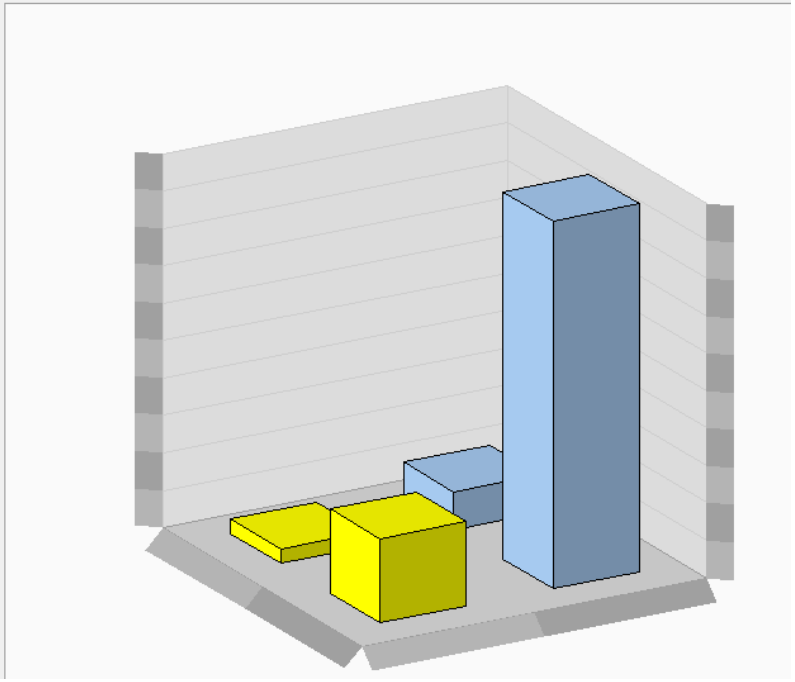
Závažnost **Z**: Accident_Severity(Serious)

Použito k hledání dodatečných podmínek - viz dále

Antecedent: Sex(Male) & Engine_Capacity__CC_(<4;1199))
Succedent: Accident_Severity(Serious)
Condition: (empty)

TEXT | DATA | FREQUENCIES | 9-FOLD TABLE | FREQDIFF | PIE | BAYES

	Succedent	¬Succedent
Ant.	16016	40520
¬Ant.	89738	392715



Dodatečné podmínky, za kterých je relativní četnost stejné závažnosti podprůměrná

- Zadání parametrů pro proceduru 4ft-Miner
- Přehled výsledků
- Nejsilnější pravidlo - detail

Zadání parametrů pro proceduru 4ft-Miner

- Minimálně 1000 nehod
- Minimálně o 10% nižší relativní četnost

ANTECEDENT

Podmínka P ke které hledáme dodatečné podmínky	Con, 2 - 2
» Engine_Capacity__CC_(<4;1199))	B, pos
» Sex_of_Driver(Male)	B, pos
Accident specifics	Con, 0 - 5
» Area_type (subset), 1 - 1	B, pos
» 1st_Road_Class (subset), 1 - 1	B, pos
» Carriageway_Hazards (subset), 1 - 1	B, pos
» Junction_Detail (subset), 1 - 1	B, pos
» Light_Conditions (subset), 1 - 1	B, pos
» Road_Surface_Conditions (subset), 1 - 1	B, pos
» Speed_limit _ed3 (subset), 1 - 1	B, pos
» Weather_Conditions (subset), 1 - 1	B, pos
Driver	Con, 0 - 5
» Age_Groups_of_Driver (subset), 1 - 1	B, pos
» Driver_Home_Area_Type (subset), 1 - 1	B, pos

Total length: 2 - 5

QUANTIFIERS

BASE	p= 1000 Abs.
BAD	p= 0.100

Generation information

Status: Solved, 16 run(s)

Mode: Standard

SUCCEDENT

Default Partial Cedent	Con, 1 - 5
» Accident_Severity(Serious)	B, pos

Total length: 1

Závažnost **Z**: Accident_Severity(Serious)

Podmínka **P**: Sex(Male) ∧ Engine_Capacity__CC_⟨4,1199)

Ostatní dílčí cedenty analogicky k předchozí úloze

List of partial cedents settings	Number of attributes	Length
Podmínka P ke které hledáme dodatečné podmínky	2	2 - 2
Accident specifics	8	0 - 5
Driver	4	0 - 5
Time	7	0 - 5
Vehicle	4	0 - 5

Přehled výsledků

Task run

Start: 15.2.2020 09:56:32 Total time: 0h 3m 56s

Number of verifications: 43588

Number of hypotheses: 18 Mode: Standard

Add group | Del group | Edit group

Actual group of hypotheses: All hypotheses

Hypotheses in group: 18 Shown hypotheses: 18 Highlighted: 0

Delete hypotheses

Nr.	Id	BAD	Hypothesis
1	16	0.175	**Engine_Capacity__CC_**(<*1199*) & **Sex**(*Male*) & **Speed_limit**(>=*50*) & **Vehicle_Type**(*Car*) & **Propulsion_Code**(*Petrol*) >÷< **Accident_Severity**(*Serious*)
2	15	0.174	**Engine_Capacity__CC_**(<*1199*) & **Sex**(*Male*) & **Speed_limit**(>=*50*) & **Vehicle_Type**(*Car*) >÷< **Accident_Severity**(*Serious*)
3	17	0.172	**Engine_Capacity__CC_**(<*1199*) & **Sex**(*Male*) & **Speed_limit**(>=*50*) & **Vehicle_Type**(*Car*) & **Left_Hand_Drive**(*No*) >÷< **Accident_Severity**(*Serious*)
4	18	0.167	**Engine_Capacity__CC_**(<*1199*) & **Sex**(*Male*) & **Speed_limit**(>=*50*) & **Rush**(*Other*) & **Vehicle_Type**(*Car*) >÷< **Accident_Severity**(*Serious*)
5	1	0.141	**Engine_Capacity__CC_**(<*1199*) & **Sex**(*Male*) & **Area**(*Rural*) & **Vehicle_Type**(*Car*) >÷< **Accident_Severity**(*Serious*)
6	2	0.141	**Engine_Capacity__CC_**(<*1199*) & **Sex**(*Male*) & **Area**(*Rural*) & **Vehicle_Type**(*Car*) & **Propulsion_Code**(*Petrol*) >÷< **Accident_Severity**(*Serious*)
7	3	0.139	**Engine_Capacity__CC_**(<*1199*) & **Sex**(*Male*) & **Area**(*Rural*) & **Vehicle_Type**(*Car*) & **Left_Hand_Drive**(*No*) >÷< **Accident_Severity**(*Serious*)
8	4	0.133	**Engine_Capacity__CC_**(<*1199*) & **Sex**(*Male*) & **Area**(*Rural*) & **Rush**(*Other*) & **Vehicle_Type**(*Car*) >÷< **Accident_Severity**(*Serious*)
9	5	0.126	**Engine_Capacity__CC_**(<*1199*) & **Sex**(*Male*) & **Junction_Detail**(*Not at junction or within 20 metres*) & **Light_Conditions**(*Daylight*) & **Vehicle_Type**(*Car*) >÷< **Accident_**
10	14	0.125	**Engine_Capacity__CC_**(<*1199*) & **Sex**(*Male*) & **Road_Surface_Conditions**(*Wet or damp*) & **Rush**(*Other*) & **Vehicle_Type**(*Car*) >÷< **Accident_Severity**(*Serious*)
11	10	0.109	**Engine_Capacity__CC_**(<*1199*) & **Sex**(*Male*) & **Light_Conditions**(*Daylight*) & **Rush**(*Other*) & **Vehicle_Type**(*Car*) >÷< **Accident_Severity**(*Serious*)
12	9	0.107	**Engine_Capacity__CC_**(<*1199*) & **Sex**(*Male*) & **Light_Conditions**(*Daylight*) & **Weekend**(*Weekdays*) & **Vehicle_Type**(*Car*) >÷< **Accident_Severity**(*Serious*)
13	11	0.107	**Engine_Capacity__CC_**(<*1199*) & **Sex**(*Male*) & **Road_Surface_Conditions**(*Wet or damp*) & **Vehicle_Type**(*Car*) >÷< **Accident_Severity**(*Serious*)
14	12	0.107	**Engine_Capacity__CC_**(<*1199*) & **Sex**(*Male*) & **Road_Surface_Conditions**(*Wet or damp*) & **Vehicle_Type**(*Car*) & **Propulsion_Code**(*Petrol*) >÷< **Accident_Severity**(
15	13	0.104	**Engine_Capacity__CC_**(<*1199*) & **Sex**(*Male*) & **Road_Surface_Conditions**(*Wet or damp*) & **Vehicle_Type**(*Car*) & **Left_Hand_Drive**(*No*) >÷< **Accident_Severity**(*Ser*
16	6	0.102	**Engine_Capacity__CC_**(<*1199*) & **Sex**(*Male*) & **Light_Conditions**(*Daylight*) & **Vehicle_Type**(*Car*) >÷< **Accident_Severity**(*Serious*)
17	7	0.102	**Engine_Capacity__CC_**(<*1199*) & **Sex**(*Male*) & **Light_Conditions**(*Daylight*) & **Vehicle_Type**(*Car*) & **Propulsion_Code**(*Petrol*) >÷< **Accident_Severity**(*Serious*)
18	8	0.101	**Engine_Capacity__CC_**(<*1199*) & **Sex**(*Male*) & **Light_Conditions**(*Daylight*) & **Vehicle_Type**(*Car*) & **Left_Hand_Drive**(*No*) >÷< **Accident_Severity**(*Serious*)

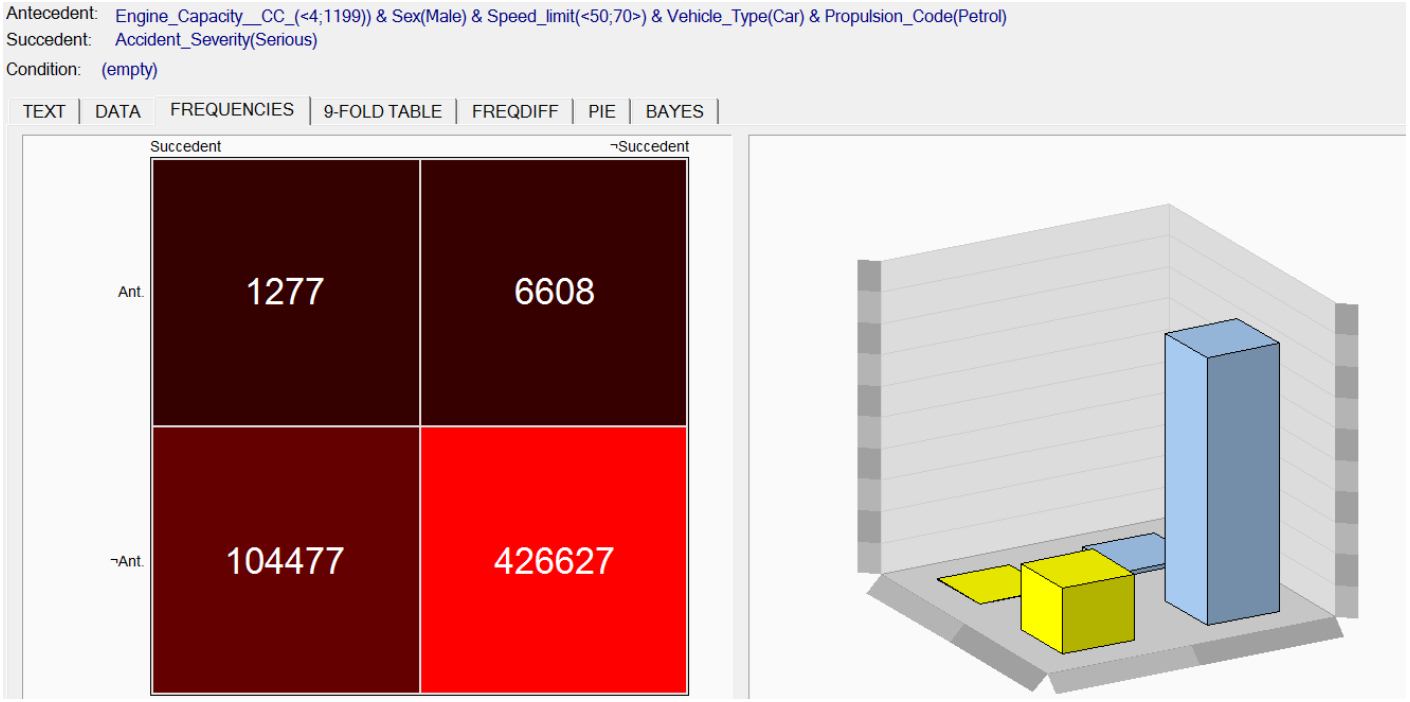
Viz příklad na začátku

Nejsilnější pravidlo - detail

Podmínka **P**: Sex(Male) ∧ Engine_Capacity__CC_⟨4,1199)

Dodatečná podmínka **D**: : Speed_limit(>=50) ∧ Vehicle_type(Car) ∧ Propulsion_Code(Petrol)

Závažnost **Z**: Accident_Severity(Serious)



$$\frac{1277}{1277+6608} = (1-0.175)\frac{1277+104477}{1277+6608+104477+426627}$$

Princip řešení

Úloha: Existuje podmínka P, dodatečná podmínka D a závažnost Z nehody takové, že

- relativní četnost nehod se závažností Z za podmínky P je vyšší než relativní četnost poruch se závažnosti Z mezi všemi nehodami
- relativní četnost nehod se závažností Z je za současné platnosti podmínky P i dodatečné D podmínky nižší, než je relativní četnost nehod se závažnosti Z mezi všemi nehodami

Postup:

1. Najdeme pravidla P ≈ Z pro která platí že relativní relativní četnost Z je nadprůměrná, je použit 4ft-kvatifikátor AAD (odpovídající liftu), viz slides 4 až 7
2. Vybereme antecedent zajímavého pravidlo jako podmínku P0, viz slide 8 (zde to nejsilnější).
3. Podmínku P0 zafixujeme jako pevnou část antecedentu a hledáme dodatečné podmínky D, je použit 4ft-kvatifikátor BAD, viz slides 10 až 12.
4. Kroky 2 a 3 můžeme opakovat pro další vhodná zajímavá pravidla.