Tato prezentace je součástí wiki-prezentace [Metoda GUHA a systém LISp-Miner]
Je dostupná z [této adresy]
Verse 11. 9. 2019

Asociační pravidla v datech s neúplnou informací - optimistické doplnění

Jan Rauch
Katedra informačního a znalostního inženýrství
Vysoká škola ekonomická

Asociační pravidla v datech s neúplnou informací - optimistické doplnění

- Matice dat s neúplnou informací
- Doplnění matice dat s neúplnou informací
- Optimistické doplnění – principy
- Devítipolní a čtyřpolní tabulka
- Optimistické doplnění pro $\varphi \Rightarrow_{0.9,30} \psi$
- Třídy 4ft-kvantifikátorů
- Optimistická doplnění pro třídy 4ft-kvantifikátorů

Matice dat s neúplnou informací

Výsledek předzpracování – matice dat $\mathcal{M}^X$ s neúplnou informací :

$\mathcal{M}^X$	A_1	A_2	A_3	$\dots$	A_P
o_1	1	X	5	$\dots$	1
o_2	X	1	X	$\dots$	3
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
o_{n-1}	2	X	2	$\dots$	2
o_n	X	3	1	$\dots$	X

X – kód chybějící hodnoty

Doplnění matice dat s neúplnou informací

$\mathcal{M}^X$	A_1	A_2	A_3	$\dots$	A_P
o_1	1	X	5	$\dots$	1
o_2	X	1	X	$\dots$	3
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
o_{n-1}	2	X	2	$\dots$	2
o_n	X	3	1	$\dots$	X

Matice dat $\mathcal{M}^X$

$\mathcal{M}$	A_1	A_2	A_3	$\dots$	A_P
o_1	1	3	5	$\dots$	1
o_2	3	1	4	$\dots$	3
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
o_{n-1}	2	2	2	$\dots$	2
o_n	2	3	1	$\dots$	1

Matice dat $\mathcal{M}$

Obrázek 4.6. Matice dat $\mathcal{M}_X$ s neúplnou informací a příklad jejího doplnění $\mathcal{M}$

Každá chybějící hodnota se nahradí některou z přípustných hodnot

Optimistické doplnění – principy

- Týká se pouze asociačních pravidel
- Vychází z trojhodnotových booleovských atributů popsaných zde.
- Asociační pravidla $\varphi \approx \psi$ v matici dat $\mathcal{M}^X$ s neúplnou informací je
  - pravdivé, tedy Val($\varphi \approx \psi$, $\mathcal{M}^X$) = 1, pokud je pravdivé alespoň v jednom doplnění matice $\mathcal{M}^X$
  - nepravdivé, tedy Val($\varphi \approx \psi$, $\mathcal{M}^X$) = 0, pokud je nepravdivé v každém doplnění matice $\mathcal{M}^X$.

Devítipolní a čtyřpolní tabulka (1)

$\mathcal{M}^X$	φ	ψ
o_1	1	1
o_2	X	1
o_3	0	X
o_4	0	0
…	…	…
o_{n-1}	X	0
o_n	1	X

$\Rightarrow$

$\mathcal{M}^X$	ψ	ψ_X	$\neg\psi$
φ	$f_{1,1}$	$f_{1,X}$	$f_{1,0}$
φ_X	$f_{X,1}$	$f_{X,X}$	$f_{X,0}$
$\neg\varphi$	$f_{0,1}$	$f_{0,X}$	$f_{0,0}$

Devítipolní a čtyřpolní tabulka (2)

$\mathcal{M}^X$	φ	ψ
o_1	1	1
o_2	X	1
o_3	0	X
o_4	0	0
…	…	…
o_{n-1}	X	0
o_n	1	X

$\mathcal{M}^X$	ψ	ψ_X	$\neg\psi$
φ	$f_{1,1}$	$f_{1,X}$	$f_{1,0}$
φ_X	$f_{X,1}$	$f_{X,X}$	$f_{X,0}$
$\neg\varphi$	$f_{0,1}$	$f_{0,X}$	$f_{0,0}$

$\mathcal{M}$	ψ	$\neg\psi$
φ	$f_{1,1} + f_{1,X,a} +$ $+ f_{X,1,a} + f_{X,X,a}$	$f_{1,0} + f_{1,X,b} +$ $+ f_{X,0,b} + f_{X,X,b}$
$\neg\varphi$	$f_{0,1} + f_{X,1,c} +$ $+ f_{0,X,c} + f_{X,X,c}$	$f_{0,0} + f_{X,0,b} +$ $+ f_{0,X,d} + f_{X,X,d}$

Optimistické doplnění pro $\varphi \Rightarrow_{0.9,30} \psi$ (1)

$$\boxed{\frac{a}{a+b} \geq 0.9 \wedge a \geq 30}$$

Nejlepší případ:

- *a* co největší
- *b* co nejmenší

$\mathcal{M}$	ψ	$\neg\psi$
φ	a	b
$\neg\varphi$	c	d

$\mathcal{M}^X$	ψ	ψ_X	$\neg\psi$
φ	$f_{1,1}$	$f_{1,X}$	$f_{1,0}$
φ_X	$f_{X,1}$	$f_{X,X}$	$f_{X,0}$
$\neg\varphi$	$f_{0,1}$	$f_{0,X}$	$f_{0,0}$

$\mathcal{M}$	ψ	$\neg\psi$
φ	$f_{1,1} + f_{1,X} + f_{X,X} + f_{X,1}$	$f_{1,0}$
$\neg\varphi$	$f_{0,1} + f_{0,X,c}$	$f_{0,0} + f_{X,0} + f_{0,X,d}$

=

$\mathcal{M}$	ψ	$\neg\psi$
φ	a_o	b_o
$\neg\varphi$	c_o	d_o

$$f_{0,X,c} + f_{0,X,d} = f_{0,X}$$

Optimistické doplnění pro $\varphi \Rightarrow_{0.9,30} \psi$ (2)

Čtyřpolní tabulka

$\mathcal{M}$	ψ	$\neg\psi$
φ	a_o	b_o
$\neg\varphi$	c_o	d_o

je optimistickým doplněním

devítipolní tabulky

$\mathcal{M}^X$	ψ	ψ_X	$\neg\psi$
φ	$f_{1,1}$	$f_{1,X}$	$f_{1,0}$
φ_X	$f_{X,1}$	$f_{X,X}$	$f_{X,0}$
$\neg\varphi$	$f_{0,1}$	$f_{0,X}$	$f_{0,0}$

pro $\Rightarrow_{0.9,30}$.

Val($\varphi \approx \psi$, $\mathcal{M}^X$) = 1: Val($\varphi \Rightarrow_{0.9,30} \psi$, $\mathcal{M}$) = 1 je pravdivé alespoň v jednom doplnění matice $\mathcal{M}^X$ a to nastane právě když $\frac{a_o}{a_o + b_o} \geq 0.9 \wedge a \geq 30$.

Třídy 4ft-kvantifikátorů

Třída 4ft-kvantifikátorů	Podmínka zachování pravdivosti
implikační kvantifikátory	$a' \geq a \wedge b' \leq b$
slabě implikační kvantifikátory	$a+b+c+d = a'+b'+'c+d' \wedge$ $\wedge\ a' \geq a\ \wedge\ b' \leq b$
dvojitě implikační kvantifikátory	$a' \geq a \wedge b' \leq b \wedge c' \leq c$
slabě dvojitě implikační kvantifikátory	$a+b+c+d = a'+b'+'c+d' \wedge$ $\wedge\ a \wedge a' \geq a \wedge b' \leq b \wedge c' \leq c$
Σ-dvojitě implikační kvantifikátory	$a' \geq a \wedge b'+c' \leq b+c$
slabě Σ-dvojitě implikační kvantifikátory	$a+b+c+d = a'+b'+'c+d' \wedge$ $\wedge\ a \wedge a' \geq a \wedge b'+c' \leq b+c$
ekvivalenční kvantifikátory	$a' \geq a \wedge b' \leq b \wedge c' \leq c \wedge d' \geq d$
slabě ekvivalenční kvantifikátory	$a+b+c+d = a'+b'+'c+d' \wedge$ $\wedge\ a' \geq a \wedge b' \leq b \wedge c' \leq c \wedge d' \geq d$
Σ-ekvivalenční kvantifikátory	$a'+d' \geq a+d \wedge b'+c' \leq b+c$
slabě Σ-ekvivalenční kvantifikátory	$a+b+c+d = a'+b'+'c+d' \wedge$ $\wedge\ a'+d' \geq a+d$
symetrické kvantifikátory	$a' = a \wedge b' = c \wedge c' = b \wedge d' = d$
s vlastností F	$(b \geq c-1 \geq 0 \wedge$ $\wedge\ a' = a \wedge b' = b+1 \wedge c' = c-1 \wedge d' = d) \vee$ $\vee\ (c \geq b-1 \geq 0 \wedge$ $\wedge\ a' = a \wedge b' = b-1 \wedge c' = c+1b \wedge d' = d)$

Další informace o třídách 4ft-kvantifikátorů jsou zde.

Optimistická doplnění pro třídy 4ft-kvantifikátorů (1)

$\mathcal{M}^X$	ψ	ψ_X	$\neg\psi$
φ	$f_{1,1}$	$f_{1,X}$	$f_{1,0}$
φ_X	$f_{X,1}$	$f_{X,X}$	$f_{X,0}$
$\neg\varphi$	$f_{0,1}$	$f_{0,X}$	$f_{0,0}$

Pokud $\approx$ je slabě implikační nebo implikační, pak každá čtyřpolní tabulka

$\mathcal{M}$	ψ	$\neg\psi$
φ	$f_{1,1} + f_{1,X} + f_{X,X} + f_{X,1}$	$f_{1,0}$
$\neg\varphi$	$f_{0,1} + f_{0,X,c}$	$f_{0,0} + f_{X,0} + f_{0,X,d}$

taková, že $f_{0,X,c} + f_{0,X,d} = f_{0,X}$ je optimistické doplnění tabulky $9ft(\varphi, \psi, \mathcal{M}^X)$ pro $\approx$. Neformálně lze říci, že frekvenci a ze čtyřpolní tabulky uděláme co největší a frekvenci b necháme stejnou.

Optimistická doplnění pro třídy 4ft-kvantifikátorů (2)

$\mathcal{M}^X$	ψ	ψ_X	$\neg\psi$
φ	$f_{1,1}$	$f_{1,X}$	$f_{1,0}$
φ_X	$f_{X,1}$	$f_{X,X}$	$f_{X,0}$
$\neg\varphi$	$f_{0,1}$	$f_{0,X}$	$f_{0,0}$

Pokud $\approx$ je slabě Σ – dvojitě implikační nebo Σ – dvojitě implikační, pak každá čtyřpolní tabulka

$\mathcal{M}$	ψ	$\neg\psi$
φ	$f_{1,1} + f_{1,X} + f_{X,X} + f_{X,1}$	$f_{1,0}$
$\neg\varphi$	$f_{0,1}$	$f_{0,0} + f_{0,X} + f_{X,0}$

je optimistické doplnění tabulky $9ft(\varphi, \psi, \mathcal{M}^X)$ pro $\approx$. Neformálně lze říci, že frekvenci a ze čtyřpolní tabulky uděláme co největší a frekvence b a c necháme stejné.

Optimistická doplnění pro třídy 4ft-kvantifikátorů (3)

$\mathcal{M}^X$	ψ	ψ_X	$\neg\psi$
φ	$f_{1,1}$	$f_{1,X}$	$f_{1,0}$
φ_X	$f_{X,1}$	$f_{X,X}$	$f_{X,0}$
$\neg\varphi$	$f_{0,1}$	$f_{0,X}$	$f_{0,0}$

Pokud $\approx$ slabě Σ – ekvivalenční nebo Σ – ekvivalenční, pak každá čtyřpolní tabulka

$\mathcal{M}$	ψ	$\neg\psi$
φ	$f_{1,1} + f_{1,X} + f_{X,X,a} + f_{X,1}$	$f_{1,0}$
$\neg\varphi$	$f_{0,1}$	$f_{0,0} + f_{0,X} + f_{X,0} + f_{X,X,d}$

taková, že $f_{X,X,a} + f_{X,X,d} = f_{X,X}$ je optimistické doplnění tabulky $9ft(\varphi, \psi, \mathcal{M}^X)$ pro $\approx$. Neformálně lze říci, že součet frekvencí $a + d$ ze čtyřpolní tabulky uděláme co největší a frekvence b a c necháme stejné.

Optimistická doplnění pro třídy 4ft-kvantifikátorů (4)

$\mathcal{M}^X$	ψ	ψ_X	$\neg\psi$
φ	$f_{1,1}$	$f_{1,X}$	$f_{1,0}$
φ_X	$f_{X,1}$	$f_{X,X}$	$f_{X,0}$
$\neg\varphi$	$f_{0,1}$	$f_{0,X}$	$f_{0,0}$

Pokud $\approx$ je symetrický 4ft-kvantifikátor s vlastností F, pak čtyřpolní tabulka

$\mathcal{M}$	ψ	$\neg\psi$
φ	$f_{1,1} + f_{1,X} + f_{X,X,a} + f_{X,1}$	$f_{1,0}$
$\neg\varphi$	$f_{0,1}$	$f_{0,0} + f_{X,0} + f_{0,X} + f_{X,X,d}$

taková, že $f_{X,X,a} + f_{X,X,d} = f_{X,X}$ a zároveň absolutní hodnota rozdílu frekvencí $|(f_{1,1} + f_{1,X} + f_{X,X,a} + f_{X,1}) - (f_{0,0} + f_{X,0} + f_{0,X} + f_{X,X,d})|$ je minimální možná, je optimistické doplnění tabulky $9ft(\varphi, \psi, \mathcal{M}^X)$ pro $\approx$.