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

Je dostupná z [této adresy]

Verse 6. 9. 2019

Asociační pravidla v datech s neúplnou informací - úvahy a příklady

prof. RNDr. Jan Rauch, CSc.

Katedra informačního a znalostního inženýrství

Asociační pravidla v datech s neúplnou informací - úvahy a příklady

- Hodnota asociačního pravidla v matici dat $\mathcal{M}^X$ s neúplnou informací
- Určení hodnoty $\varphi \approx \psi$ v matici dat $\mathcal{M}^X$ pomocí definice
- Devítipolní a čtyřpolní tabulka
- Určení hodnoty $\varphi \approx \psi$ v matici dat $\mathcal{M}^X$ pomocí devítipolní a čtyřpolních tabulek
- Zabezpečené doplnění pro $\varphi \Rightarrow_{0.9,30} \psi$
- Zabezpečené doplnění pro implikační kvantifikátor $\approx$
- Zabezpečené doplnění pro $\varphi \sim^{+}_{0.5,30} \psi$
- Zabezpečené doplnění – poznámky

Hodnota asociačního pravidla v matici dat $\mathcal{M}^X$ s neúplnou informací

- Val($\varphi \approx \psi$, $\mathcal{M}^X$) = 1
  Asociační pravidlo $\varphi \approx \psi$ je pravdivé v matici dat $\mathcal{M}^X$ s neúplnou informací, pokud je pravdivé ve všech možných doplněních matice $\mathcal{M}^X$.

- Val($\varphi \approx \psi$, $\mathcal{M}^X$) = 0
  Asociační pravidlo je nepravdivé v matici dat $\mathcal{M}^X$ s neúplnou informací, pokud je nepravdivé ve všech možných doplněních matice $\mathcal{M}^X$.

- Val($\varphi \approx \psi$, $\mathcal{M}^X$) = X
  Hodnotou asociačního pravidla je X jindy, pokud existuje doplnění $\mathcal{M}_1$ matice dat $\mathcal{M}^X$ ve kterém je pravidlo pravdivé a doplnění $\mathcal{M}_2$ matice dat $\mathcal{M}^X$, ve kterém je pravidlo nepravdivé.

Určení hodnoty $\varphi \approx \psi$ v matici dat $\mathcal{M}^X$ pomocí definice

- $Val(\varphi \approx \psi, \mathcal{M}^X) = 1$: $Val(\varphi \approx \psi, \mathcal{M}) = 1$ pro každé doplnění $\mathcal{M}$ matice $\mathcal{M}^X$
- $Val(\varphi \approx \psi, \mathcal{M}^X) = 0$: $Val(\varphi \approx \psi, \mathcal{M}) = 0$ pro každé doplnění $\mathcal{M}$ matice $\mathcal{M}^X$
- $Val(\varphi \approx \psi, \mathcal{M}^X) = X$: jindy

$\mathcal{M}^X$	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

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

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

$\approx(a,b,c,d)$

1

0

Problém – mnoho doplnění

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

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

Určení hodnoty $\varphi \approx \psi$ v matici dat $\mathcal{M}^X$ pomocí devítipolní a čtyřpolních tabulek

- Val($\varphi \approx \psi$, $\mathcal{M}^X$) = 1: Val($\varphi \approx \psi$, $\mathcal{M}$) = 1 pro každé doplnění $\mathcal{M}$ matice $\mathcal{M}^X$
- Val($\varphi \approx \psi$, $\mathcal{M}^X$) = 0 : Val($\varphi \approx \psi$, $\mathcal{M}$) = 0 pro každé doplnění $\mathcal{M}$ matice $\mathcal{M}^X$
- Val($\varphi \approx \psi$, $\mathcal{M}^X$) = X : jindy

$\mathcal{M}^X$	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	a	b
$\neg\varphi$	c	d

⇒ $\approx$(a,b,c,d) → 1 / 0

Problém – mnoho tabulek $\langle$a,b,c,d$\rangle$

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

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

Nejhorší případ:

- a co nejmenší
- b co největší

$\mathcal{M}$	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	$f_{X,1}$	$f_{X,X}$	$f_{X,0}$
$\neg\varphi$	$f_{0,1}$	$f_{0,X}$	$f_{0,0}$

$\Downarrow$

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

=

$\mathcal{M}$	ψ	$\neg\psi$
φ	a_z	b_z
$\neg\varphi$	c_z	d_z

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

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

Čtyřpolní tabulka

$\mathcal{M}$	ψ	$\neg\psi$
φ	a_z	b_z
$\neg\varphi$	c_z	d_z

je zabezpečeným 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 pro každé doplnění $\mathcal{M}$ matice $\mathcal{M}^X$

a to nastane právě když $\frac{a_z}{a_z + b_z} \geq 0.9 \wedge a \geq 30$.

Zabezpečené doplnění pro implikační kvantifikátor ≈

Když ≈(a,b,c,d) = 1
a $a' \geq a \wedge b' \leq b$,
pak i ≈(a',b',c',d') = 1

Nejhorší případ:

- *a* co nejmenší
- *b* co největší

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

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

Zabezpečené doplnění pro φ $\sim^{+}_{0.5,30}$ ψ (1)

$$\boxed{\frac{a}{a+b} \geq (1+0.5)\frac{a+c}{a+b+c+d} \wedge a \geq 30}$$

$\mathcal{M}$	ψ	¬ψ
φ	a	b
¬φ	c	d

Nejhorší případ: ?

- a co nejmenší kvůli $a \geq 30$
- …?

$$\frac{a(a+b+c+d)}{(a+b)(a+c)} \geq (1+0.5)$$

$$\frac{\boxed{a^2+ab+ac}+ad}{\boxed{a^2+ab+ac}+bc} \geq (1+0.5)$$

Nejhorší případ:

- a co nejmenší kvůli $a \geq 30$
- ad co nejmenší a bc co největší

  kvůli co nejmenší hodnotě $\dfrac{a^2+ab+ac+ad}{a^2+ab+ac+bc}$.

Zabezpečené doplnění pro $\varphi \sim^{+}_{0.5,30} \psi$ (2)

Příklad:

$\mathcal{M}^X$	ψ	ψ_X	$\neg\psi$
φ	11	0	0
φ_X	0	20	0
$\neg\varphi$	0	0	11

$\Rightarrow$

$\mathcal{M}$	ψ	$\neg\psi$
φ	11	10
$\neg\varphi$	10	11

10x10 = 100
9x11 = 99
8x12 = 96
…
1x19 = 19
0x20 = 0

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

$\Rightarrow$

$\mathcal{M}$	ψ	$\neg\psi$
φ	$f_{1,1}$	b_z
$\neg\varphi$	c_z	$f_{0,0}$

$\boxed{|b_z - c_z| \text{ co nejmenší}}$

Zabezpečené doplnění – poznámky

- Zabezpečené doplnění pro 4ft-kvantifikátor $\approx$ záleží na třídě 4ft-kvantifikátorů, do které $\approx$ patří.
- Fisherův kvantifikátor i χ^2 – kvantifikátor patří do stejné třídy jako $\sim^{+}_{p,Base}$ a mají také stejné zabezpečené doplnění.

Fisherův kvantifikátor: $\sum_{i=a}^{\min(r,k)} \frac{\binom{k}{i}\binom{n-k}{r-i}}{\binom{n}{r}} \leq \alpha \;\wedge\; ad > bc$

$$r = a + b,\; k = a + c \text{ a } n = a + b + c + d$$

χ^2 – kvantifikátor: $\frac{(ad-bc)^2}{rk(n-k)(n-r)} n \;\geq\; \chi^2_\alpha \;\wedge\; ad \;>\; bc$