Miary oceny reguł asocjacyjnych

Marzena Kryszkiewicz
Politechnika Warszawska

Relatywne wsparcie reguły asocjacyjnej

- *Relatywne wsparcie* reguły asocjacyjnej $X \rightarrow Y$ jest definiowane jako prawdopodobieństwo współwystąpienia X i Y:

$$rSup(X \rightarrow Y) = P(XY) = sup(XY) / |D|.$$

- **Własność.** *Relatywne wsparcie* zależy od prawdopodobieństwa wystąpienia bazy reguły, ale nie zależy od prawdopodobieństwa wystąpienia jej poprzednika, ani następnika.

2

Zaufanie reguły asocjacyjnej

- *Zaufanie* reguły asocjacyjnej $X \rightarrow Y$ jest definiowane jako warunkowe prawdopodobieństwo wystąpienia Y pod warunkiem wystąpienia X:

$$conf(X \rightarrow Y) = P(XY) / P(X).$$

- **Własność.** *Zaufanie* zależy od prawdopodobieństw wystąpienia poprzednika i bazy reguły, ale nie zależy od prawdopodobieństwa wystąpienia jej następnika.

3

(Nie)zależność zdarzeń

- X i Y są *niezależne*, jeśli:
  $$P(XY) = P(X) \times P(Y).$$
- Wpp. są zwane *zależnymi*.
- **Naturalna interpretacja:**
  - Jeżeli $P(XY) > P(X) * P(Y)$, to X i Y są zależne *pozytywnie*.
  - Jeżeli $P(XY) = P(X) * P(Y)$, to X i Y są niezależne.
  - Jeżeli $P(XY) < P(X) * P(Y)$, to X i Y są zależne *negatywnie*.
- **Własność.** Fakt, że $P(XY) \neq P(X) * P(Y)$ nie określa stopnia, w jakim X i Y są zależne.

4

Zaufanie a (nie)zależność zdarzeń…

TId	X	Y
1		x
2		x
3		x
4		x
5		x
6		x
7		x
8		x
9	x	x
10	x	

- $P(X) = \frac{2}{10}$, $P(Y) = \frac{9}{10} \Rightarrow P(X) \times P(Y) = 0.18$.
- $P(XY) = \frac{1}{10} = 0.1 < 0.18 = P(X) \times P(Y)$, czyli **$X$ i Y są zależne negatywnie.**
- ***$conf(X \rightarrow Y) = P(XY) / P(X) = \frac{1}{2}$.***

- $P(X) = \frac{2}{10}$, $P(\bar{Y}) = \frac{1}{10} \Rightarrow P(X) \times P(\bar{Y}) = 0.02$.
- $P(X\bar{Y}) = \frac{1}{10} = 0.1 > 0.02 = P(X) \times P(\bar{Y})$, czyli **$X$ i $\bar{Y}$ są zależne pozytywnie.**
- ***$conf(X \rightarrow \bar{Y}) = P(X\bar{Y}) / P(X) = \frac{1}{2}$.***

5

Zaufanie a (nie)zależność zdarzeń

TId	X	Y
1		x
2		x
3		x
4		x
5		x
6		x
7		x
8	x	x
9	x	x
10	x	

- $P(X) = \frac{3}{10}$, $P(Y) = \frac{9}{10} \Rightarrow P(X) \times P(Y) = 0.27$.
- $P(XY) = \frac{2}{10} = 0.2 < 0.27 = P(X) \times P(Y)$, czyli **$X$ i Y są zależne negatywnie.**
- ***$conf(X \rightarrow Y) = P(XY) / P(X) = \frac{2}{3}$.***

- $P(X) = \frac{3}{10}$, $P(\bar{Y}) = \frac{1}{10} \Rightarrow P(X) \times P(\bar{Y}) = 0.03$.
- $P(X\bar{Y}) = \frac{1}{10} = 0.1 > 0.03 = P(X) \times P(\bar{Y})$, czyli **$X$ i $\bar{Y}$ są zależne pozytywnie.**
- ***$conf(X \rightarrow \bar{Y}) = P(X\bar{Y}) / P(X) = \frac{1}{3}$.***

6

Współczynnik podniesienia (lift)

- *Współczynnik podniesienia* (ang. *lift*) dla reguły asocjacyjnej $X \rightarrow Y$ jest definiowany jako stosunek prawdopodobieństwa warunkowego wystąpienia Y pod warunkiem wystąpienia X do prawdopodobieństwa wystąpienia Y:

$$lift(X \rightarrow Y) = \frac{conf(X \rightarrow Y)}{P(Y)} = \frac{P(XY)}{P(X) \times P(Y)}.$$

- **Własność.**

(Nie)zależność	Warunek na (nie)zależność	Równoważny warunek	Równoważny warunek
Y i X są zależne pozytywnie	$P(XY) > P(X) \times P(Y)$	$lift(X \rightarrow Y) > 1$	$lift(Y \rightarrow X) > 1$
Y i X są niezależne	$P(XY) = P(X) \times P(Y)$	$lift(X \rightarrow Y) = 1$	$lift(Y \rightarrow X) = 1$
Y i X są zależne negatywnie	$P(XY) < P(X) \times P(Y)$	$lift(X \rightarrow Y) < 1$	$lift(Y \rightarrow X) < 1$

- **Własność.** $lift(X \rightarrow Y) = lift(Y \rightarrow X)$.

7

Lift a (nie)zależność zdarzeń

TId	X	Y	Z	V
1	x	x		
2	x	x		
3	x	x		
4	x	x		
5	x	x		
6	x	x		
7	x	x		
8	x	x		
9	x	x		
10			x	x

- $P(X) = \frac{9}{10}$, $P(Y) = \frac{9}{10} \Rightarrow P(X) \times P(Y) = 0.81$.
- $P(XY) = \frac{9}{10} = 0.9 > 0.81 = P(X) \times P(Y)$, czyli **$X$ i Y są zależne pozytywnie.**
- $\boldsymbol{lift(X \rightarrow Y) = \frac{P(XY)}{P(X) \times P(Y)} = \frac{0.9}{0.9 \times 0.9} = \frac{10}{9} = 1\frac{1}{9}}$.

- $P(Z) = \frac{1}{10}$, $P(V) = \frac{1}{10} \Rightarrow P(Z) \times P(V) = 0.01$.
- $P(ZV) = \frac{1}{10} = 0.1 > 0.01 = P(Z) \times P(V)$, czyli **$Z$ i V są zależne pozytywnie.**
- $\boldsymbol{lift(Z \rightarrow V) = \frac{P(ZV)}{P(Z) \times P(V)} = \frac{0.1}{0.1 \times 0.1} = \frac{10}{1} = 10}$.

8

Lift a (nie)zależność zdarzeń

TId	X	Y	Z	V
1	x	x		
2	x	x		
3	x	x		
4	x	x		
5	x	x		
6	x	x		
7	x	x		
8	x	x		x
9	x	x	x	
10			x	x

- $P(X) = \frac{9}{10}$, $P(Y) = \frac{9}{10} \Rightarrow P(X) \times P(Y) = 0.81$.
- $P(XY) = \frac{9}{10} = 0.9 > 0.81 = P(X) \times P(Y)$, czyli **$X$ i Y są zależne pozytywnie.**
- $\boldsymbol{lift(X \rightarrow Y) = \frac{P(XY)}{P(X) \times P(Y)} = \frac{0.9}{0.9 \times 0.9} = \frac{10}{9} = 1\frac{1}{9}}$.

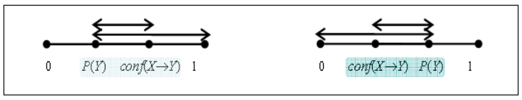
- $P(Z) = \frac{2}{10}$, $P(V) = \frac{2}{10} \Rightarrow P(Z) \times P(V) = 0.04$.
- $P(ZV) = \frac{1}{10} = 0.1 > 0.04 = P(Z) \times P(V)$, czyli **$Z$ i V są zależne pozytywnie.**
- $\boldsymbol{lift(Z \rightarrow V) = \frac{P(ZV)}{P(Z) \times P(V)} = \frac{0.1}{0.2 \times 0.2} = \frac{10}{4} = 2.5}$.

9

Współczynnik pewności (cf)…

- Współczynnik pewności (ang. *certainty factor*, *cf*) dla reguły asocjacyjnej $X \rightarrow Y$ jest definiowany jako stopień, w jakim prawdopodobieństwo wystąpienia Y może ulec zmianie pod warunkiem wystąpienia X:

$$cf(X \rightarrow Y) = \begin{cases} \frac{conf(X \rightarrow Y) - P(Y)}{1 - P(Y)} & \text{if } conf(X \rightarrow Y) > P(Y), \\ 0 & \text{if } conf(X \rightarrow Y) = P(Y), \\ -\frac{P(Y) - conf(X \rightarrow Y)}{P(Y) - 0} & \text{if } conf(X \rightarrow Y) < P(Y). \end{cases}$$



10

Współczynnik pewności (cf)

$$cf(X \rightarrow Y) = \begin{cases} \frac{conf(X \rightarrow Y) - P(Y)}{1 - P(Y)} & \text{if } conf(X \rightarrow Y) > P(Y), \\ 0 & \text{if } conf(X \rightarrow Y) = P(Y), \\ -\frac{P(Y) - conf(X \rightarrow Y)}{P(Y) - 0} & \text{if } conf(X \rightarrow Y) < P(Y). \end{cases}$$

- **Własność:**

asymetryczne dla X i Y

$$cf(X \rightarrow Y) = \begin{cases} \frac{P(XY) - P(X) \times P(Y)}{P(X) - P(X) \times P(Y)} & \text{if } P(XY) > P(X) \times P(Y), \\ 0 & \text{if } P(XY) = P(X) \times P(Y), \\ -\frac{P(X) \times P(Y) - P(XY)}{P(X) \times P(Y) - 0} & \text{if } P(XY) < P(X) \times P(Y). \end{cases}$$

symetryczne dla X i Y

11

Wartości ekstremalne miar regułowych

miara	maks.	min.
$P(XY)$	1	0
$conf(X \rightarrow Y)$	1	0
$lift(X \rightarrow Y)$	∞	0
$cf(X \rightarrow Y)$	1 dla zależności pozytywnej pomiędzy X i Y	-1 dla zależności negatywnej pomiędzy X i Y

12

Wartości ekstremalne *sup*(XY) dla zadanych wartości *sup*(X) i *sup*(Y)

- Własność.

a) $max\_sup(XY|_{sup(X),\, sup(Y)}) = \min\{sup(X), sup(Y)\}$

b) $min\_sup(XY|_{sup(X),\, sup(Y)}) =$

$$\begin{cases} 0 & \text{gdy } sup(X) + sup(Y) \le |D| \\ sup(X) + sup(Y) - |D| & \text{gdy } sup(X) + sup(Y) > |D| \end{cases}$$

X	Y
x	x
x	x
x	

X	Y
x	
x	
x	
	x
	x

X	Y
x	
x	
x	x
x	x
x	x
	x

Uwaga: $sup(X \vee Y) = sup(X) + sup(Y) - sup(XY)$

13

Wartości ekstremalne P(XY) dla zadanych wartości P(X) i P(Y)

- Własność.

a) $max\_P(XY|_{P(X),\, P(Y)}) = \min\{P(X), P(Y)\}$

b) $$min\_P(XY|_{P(X),\, P(Y)}) = \begin{cases} 0 & \text{if } P(X) + P(Y) \le 1 \\ P(X) + P(Y) - 1 & \text{if } P(X) + P(Y) > 1 \end{cases}$$

$= \max\{0,\ P(X) + P(Y) - 1\}$

X	Y
x	x
x	x
x	

X	Y
x	
x	
x	
	x
	x

X	Y
x	
x	
x	x
x	x
x	x
	x

Uwaga: $P(X \vee Y) = P(X) + P(Y) - P(XY)$

14

Wartości ekstremalne miar dla zadanych wartości P(X) i P(Y)

- Własności.

zamiast 1   zamiast 0

miara	maks. dla ustalonych wartości $P(X)$ i $P(Y)$	min. dla ustalonych wartości $P(X)$ i $P(Y)$
$P(XY)$	$\min\{P(X), P(Y)\}$	$\max\{0,\ P(X) + P(Y) - 1\}$
$conf(X \rightarrow Y)$	$\frac{\min\{P(X), P(Y)\}}{P(X)}$	$\frac{\max\{0, P(X) + P(Y) - 1\}}{P(X)}$
$lift(X \rightarrow Y)$	$\frac{\min\{P(X), P(Y)\}}{P(X) \times P(Y)}$	$\frac{\max\{0, P(X) + P(Y) - 1\}}{P(X) \times P(Y)}$
$cf(X \rightarrow Y)$	$\frac{\min\{P(X), P(Y)\} - P(X) \times P(Y)}{P(X) - P(X) \times P(Y)}$ dla zależności pozytywnej pomiędzy X i Y	$-\frac{P(X) \times P(Y) - \max\{0, P(X) + P(Y) - 1\}}{P(X) \times P(Y) - 0}$ dla zależności negatywnej pomiędzy X i Y

15

Współczynnik zależności (df)

- *Współczynnik zależności* (ang. *dependence factor, df*) dla reguły asocjacyjnej $X \rightarrow Y$ jest definiowany jako stopień, w jakim prawdopodobieństwo wystąpienia Y może ulec zmianie pod warunkiem wystąpienia X i Y z prawdopodobieństwem odpowiednio $P(X)$ i $P(Y)$:

zamiast 1

$$df(X \rightarrow Y) = \begin{cases} \frac{conf(X \rightarrow Y) - P(Y)}{max\_conf(X \rightarrow Y|_{P(X),P(Y)}) - P(Y)} & \text{if } conf(X \rightarrow Y) > P(Y), \\ 0 & \text{if } conf(X \rightarrow Y) = P(Y), \\ -\frac{P(Y) - conf(X \rightarrow Y)}{P(Y) - min\_conf(X \rightarrow Y|_{P(X),P(Y)})} & \text{if } conf(X \rightarrow Y) < P(Y). \end{cases}$$

zamiast 0

16

Ekstremalne wartości współczynnika zależności df

Własność. Dla dowolnych wartości $P(X)$ i $P(Y)$:

- minimalna osiągalna wartość $df(X \rightarrow Y)$ wynosi -1,
- maksymalna osiągalna wartość $df(X \rightarrow Y)$ wynosi 1.

zamiast 1

$$df(X \rightarrow Y) = \begin{cases} \frac{conf(X \rightarrow Y) - P(Y)}{max\_conf(X \rightarrow Y|_{P(X),P(Y)}) - P(Y)} & \text{if } conf(X \rightarrow Y) > P(Y), \\ 0 & \text{if } conf(X \rightarrow Y) = P(Y), \\ -\frac{P(Y) - conf(X \rightarrow Y)}{P(Y) - min\_conf(X \rightarrow Y|_{P(X),P(Y)})} & \text{if } conf(X \rightarrow Y) < P(Y). \end{cases}$$

zamiast 0

17

Współczynnik zależności df a (nie)zależność zdarzeń

TId	X	Y
1		x
2		x
3		x
4		x
5		x
6		x
7		x
8	x	x
9	x	x
10	x	

- $P(X) = \frac{3}{10}$, $P(Y) = \frac{9}{10} \Rightarrow P(X) \times P(Y) = 0.27$, $P(XY) = \frac{2}{10} = 0.2$.
- ***conf*(X → Y) = P(XY) / P(X)** $= \frac{2}{3}$.
- ***cf*(X → Y)** $= -\frac{0.9 - \frac{2}{3}}{0.9 - 0} \approx -0.26$.
- ***df*(X → Y)** $= -\frac{0.9 - \frac{2}{3}}{0.9 - \frac{2}{3}} = -1$.

===

- $P(X) = \frac{3}{10}$, $P(\bar{Y}) = \frac{1}{10} \Rightarrow P(X) \times P(\bar{Y}) = 0.03$, $P(X\bar{Y}) = \frac{1}{10} = 0.1$.
- $conf(X \rightarrow \bar{Y}) = P(X\bar{Y}) / P(X) = \frac{1}{3}$.
- $cf(X \rightarrow \bar{Y}) = \frac{\frac{1}{3} - 0.1}{1 - 0.1} \approx 0.26$.
- $df(X \rightarrow \bar{Y}) = \frac{\frac{1}{3} - 0.1}{\frac{1}{3} - 0.1} = 1$.

18

Współczynnik zależności df w terminach prawdopodobieństw…

$$df(X \to Y) = \begin{cases} \dfrac{conf(X \to Y) - P(Y)}{max\_conf(X \to Y \mid_{P(X),P(Y)}) - P(Y)} & \text{if } conf(X \to Y) > P(Y), \\ 0 & \text{if } conf(X \to Y) = P(Y), \\ -\dfrac{P(Y) - conf(X \to Y)}{P(Y) - min\_conf(X \to Y \mid_{P(X),P(Y)})} & \text{if } conf(X \to Y) < P(Y). \end{cases}$$

zamiast 1

$$df(X \to Y) = \begin{cases} \dfrac{P(XY) - P(X) \times P(Y)}{max\_P(XY \mid_{P(X),P(Y)}) - P(X) \times P(Y)} & \text{if } P(XY) > P(X) \times P(Y), \\ 0 & \text{if } P(XY) = P(X) \times P(Y), \\ -\dfrac{P(X) \times P(Y) - P(XY)}{P(X) \times P(Y) - min\_P(XY \mid_{P(X),P(Y)})} & \text{if } P(XY) < P(X) \times P(Y). \end{cases}$$

zamiast 0

19

Współczynnik zależności df w terminach prawdopodobieństw

zamiast 1

$$df(X \to Y) = \begin{cases} \dfrac{P(XY) - P(X) \times P(Y)}{max\_P(XY \mid_{P(X),P(Y)}) - P(X) \times P(Y)} & \text{if } P(XY) > P(X) \times P(Y), \\ 0 & \text{if } P(XY) = P(X) \times P(Y), \\ -\dfrac{P(X) \times P(Y) - P(XY)}{P(X) \times P(Y) - min\_P(XY \mid_{P(X),P(Y)})} & \text{if } P(XY) < P(X) \times P(Y). \end{cases}$$

zamiast 0

Własności.

- $max\_P(XY) \mid_{P(X),P(Y)}) = min\{P(X), P(Y)\}$.
- $min\_P(XY) \mid_{P(X),P(Y)}) = max\{0, P(X) + P(Y) - 1\}$.
- $df(X \to Y) = df(Y \to X)$.

20

Współczynnik zależności df a współczynnik pewności cf

- **Własność.**
  - $df(X \to Y) = cf(X \to Y) = 0$, jeśli $P(XY) = P(X) \times P(Y)$.
  - $|df(X \to Y)| \geq |cf(X \to Y)|$, jeśli $P(XY) \neq P(X) \times P(Y)$.
  - Jeśli $P(XY) > P(X) \times P(Y)$ (czyli, gdy X i Y są zależne pozytywnie), to: $df(X \to Y) = \max\{cf(X \to Y), cf(Y \to X)\}$.
  - Jeśli $P(XY) < P(X) \times P(Y)$ (czyli, gdy X i Y są zależne negatywnie), to:

    $df(X \to Y) = cf(X \to Y)$, jeśli $P(X) + P(Y) \leq 1$.

    $|df(X \to Y)| > |cf(X \to Y)|$, jeśli $P(X) + P(Y) > 1$.

21

Współczynnik zależności (df) a miara lift

$$df(X \to Y) = \begin{cases} \dfrac{conf(X \to Y) - P(Y)}{max\_conf(X \to Y \mid_{P(X),P(Y)}) - P(Y)} & \text{if } conf(X \to Y) > P(Y), \\ 0 & \text{if } conf(X \to Y) = P(Y), \\ -\dfrac{P(Y) - conf(X \to Y)}{P(Y) - min\_conf(X \to Y \mid_{P(X),P(Y)})} & \text{if } conf(X \to Y) < P(Y). \end{cases}$$

zamiast $\propto$

$$df(X \to Y) = \begin{cases} \dfrac{lift(X \to Y) - 1}{max\_lift(X \to Y \mid_{P(X),P(Y)}) - 1} & \text{if } lift(X \to Y) > 1, \\ 0 & \text{if } lift(X \to Y) = 1, \\ -\dfrac{1 - lift(X \to Y)}{1 - min\_lift(X \to Y \mid_{P(X),P(Y)})} & \text{if } lift(X \to Y) < 1. \end{cases}$$

zamiast 0

22

Przykład: Współczynnik zależności (df) a miara lift

TId	X	Y	Z	V
1	x	x		
2	x	x		
3	x	x		
4	x	x		
5	x	x		
6	x	x		
7	x	x		
8	x	x		x
9	x	x	x	
10			x	x

- $lift(X \to Y) = \dfrac{P(XY)}{P(X) \times P(Y)} = \dfrac{0.9}{0.9 \times 0.9} = \frac{10}{9} = 1\frac{1}{9}$.
- $df(X \to Y) = \dfrac{1\frac{1}{9} - 1}{\frac{0.9}{0.9 \times 0.9} - 1} = 1$.

==

- $lift(Z \to V) = \dfrac{P(ZV)}{P(Z) \times P(V)} = \dfrac{0.1}{0.2 \times 0.2} = \frac{10}{4} = 2\frac{1}{2}$.
- $df(Z \to V) = \dfrac{2\frac{1}{2} - 1}{\frac{0.2}{0.2 \times 0.2} - 1} = \dfrac{2\frac{1}{2} - 1}{5 - 1} = 0.375$.

23

Zależność pomiędzy X i Y, a zależności pomiędzy $\bar{X}$ i Y, X i $\bar{Y}$, $\bar{X}$ i $\bar{Y}$

- **Twierdzenie.**

$$df(X \to Y) = df(\bar{X} \to \bar{Y}) = -df(X \to \bar{Y}) = -df(\bar{X} \to Y).$$

24

Iloraz współczynników podniesienia (lift-ratio) i miara wzrostu (growth)

- *Iloraz współczynników podniesienia* dla $X \rightarrow Y$ jest definiowany jako:

$$lift\text{-}ratio(X \rightarrow Y) = \frac{lift(X \rightarrow Y)}{lift(X \rightarrow \bar{Y})}.$$

- *Wzrost* dla $X \rightarrow Y$ jest definiowany jako:

$$growth(X \rightarrow Y) = \frac{P(XY)}{P(Y)} : \frac{P(X\bar{Y})}{P(\bar{Y})} = \frac{sup(XY)}{sup(Y)} : \frac{sup(X\bar{Y})}{sup(\bar{Y})}.$$

- **Twierdzenie.** $lift\text{-}ratio(X \rightarrow Y) = growth(X \rightarrow Y)$.

Dowód. $lift\text{-}ratio(X \rightarrow Y) = \frac{lift(X \rightarrow Y)}{lift(X \rightarrow \bar{Y})}$

$$= \frac{P(XY)}{P(X) \times P(Y)} : \frac{P(X\bar{Y})}{P(X) \times P(\bar{Y})} = growth(X \rightarrow Y).$$

- **Przykład.** Niech $\frac{P(XY)}{P(Y)} = \frac{4}{5}$, a $\frac{P(X\bar{Y})}{P(\bar{Y})} = \frac{2}{3}$. Wtedy, $growth(X \rightarrow Y) = 1.2$.

X	Y
x	x
x	x
x	x
x	x
	x
x	
x	

25

Przykład: Lift-ratio a (nie)zależność zdarzeń

TId	X	Y	Z	V
1	x	x		
2	x	x		
3	x	x		
4	x	x		
5	x	x		
6	x	x		
7	x	x		
8	x	x		x
9	x	x	x	
10			x	x

- $lift(X \rightarrow Y) = \frac{P(XY)}{P(X) \times P(Y)} = \frac{0.9}{0.9 \times 0.9} = \frac{10}{9} = 1\frac{1}{9}$.
- $df(X \rightarrow Y) = \frac{1\frac{1}{9} - 1}{\frac{0.9}{0.9 \times 0.9} - 1} = 1$.
- $lift\text{-}ratio(X \rightarrow Y) = growth(X \rightarrow Y) = \frac{sup(XY)}{sup(Y)} : \frac{sup(X\bar{Y})}{sup(\bar{Y})} = \frac{9}{9} : \frac{0}{1} = \propto$.

====================================

- $lift(Z \rightarrow V) = \frac{P(ZV)}{P(Z) \times P(V)} = \frac{0.1}{0.2 \times 0.2} = \frac{10}{4} = 2\frac{1}{2}$.
- $df(Z \rightarrow V) = \frac{2\frac{1}{2} - 1}{\frac{0.2}{0.2 \times 0.2} - 1} = \frac{2\frac{1}{2} - 1}{5 - 1} = 0.375$.
- $lift\text{-}ratio(Z \rightarrow V) = growth(Z \rightarrow V) = \frac{sup(ZV)}{sup(V)} : \frac{sup(Z\bar{V})}{sup(\bar{V})} = \frac{1}{2} : \frac{1}{8} = 4$.

26

Iloraz szans (odds-ratio)

- *Iloraz szans* dla $X \rightarrow Y$ jest definiowany następująco:

$$odds\text{-}ratio(X \rightarrow Y) = \frac{P(XY)}{P(\bar{X}Y)} : \frac{P(X\bar{Y})}{P(\bar{X}\bar{Y})} = \frac{sup(XY)}{sup(\bar{X}Y)} : \frac{sup(X\bar{Y})}{sup(\bar{X}\bar{Y})}.$$

- **Własność.**

$$odds\text{-}ratio(X \rightarrow Y) = \frac{P(XY) \times P(\bar{X}\bar{Y})}{P(\bar{X}Y) \times P(X\bar{Y})} = \frac{sup(XY) \times sup(\bar{X}\bar{Y})}{sup(\bar{X}Y) \times sup(X\bar{Y})}.$$

X	Y
x	x
x	x
x	x
x	x
	x
x	
x	

Przykład. Niech $\frac{P(XY)}{P(\bar{X}Y)} = \frac{4}{1}$, a $\frac{P(X\bar{Y})}{P(\bar{X}\bar{Y})} = \frac{2}{1}$. Wtedy, $odds\text{-}ratio(X \rightarrow Y) = 2$.

27

Iloraz szans a (nie)zależność zdarzeń

TId	X	Y	Z	V
1	x	x		
2	x	x		
3	x	x		
4	x	x		
5	x	x		
6	x	x		
7	x	x		
8	x	x		x
9	x	x	x	
10			x	x

- $lift(X \rightarrow Y) = \frac{P(XY)}{P(X) \times P(Y)} = \frac{0.9}{0.9 \times 0.9} = \frac{10}{9} = 1\frac{1}{9}$.
- $df(X \rightarrow Y) = \frac{1\frac{1}{9} - 1}{\frac{0.9}{0.9 \times 0.9} - 1} = 1$.
- $lift\text{-}ratio(X \rightarrow Y) = growth(X \rightarrow Y) = \frac{sup(XY)}{sup(Y)} / \frac{sup(X\bar{Y})}{sup(\bar{Y})} = \propto$.
- $odds\text{-}ratio(X \rightarrow Y) = \frac{sup(XY) \times sup(\bar{X}\bar{Y})}{sup(\bar{X}Y) \times sup(X\bar{Y})} = \frac{9 \times 1}{0 \times 0} = \propto$.

==

- $lift(Z \rightarrow V) = \frac{P(ZV)}{P(Z) \times P(V)} = \frac{0.1}{0.2 \times 0.2} = \frac{10}{4} = 2\frac{1}{2}$.
- $df(Z \rightarrow V) = \frac{2\frac{1}{2} - 1}{\frac{0.2}{0.2 \times 0.2} - 1} = \frac{2\frac{1}{2} - 1}{5 - 1} = 0.375$.
- $lift\text{-}ratio(Z \rightarrow V) = growth(Z \rightarrow V) = \frac{sup(ZV)}{sup(V)} / \frac{sup(Z\bar{V})}{sup(\bar{V})} = 4$.
- $odds\text{-}ratio(Z \rightarrow V) = \frac{sup(ZV) \times sup(\bar{Z}\bar{V})}{sup(\bar{Z}V) \times sup(Z\bar{V})} = \frac{1 \times 7}{1 \times 1} = 7$.

28

Literatura...

- Sergey Brin, Rajeev Motwani, Craig Silverstein: Beyond Market Baskets: Generalizing Association Rules to Correlations. SIGMOD Conference 1997: 265-276
- Marzena Kryszkiewicz: Dependence Factor for Association Rules. ACIIDS (2) 2015: 135-145
- Marzena Kryszkiewicz: Dependence Factor as a Rule Evaluation Measure. Challenges in Computational Statistics and Data Mining 2016: 205-223
- Marzena Kryszkiewicz: Virtual Balancing of Decision Classes. ACIIDS (1) 2017: 673-684
- Marzena Kryszkiewicz: ACBC-Adequate Association and Decision Rules Versus Key Generators and Rough Sets Approximations. Fundam. Inform. 148(1-2): 65-85 (2016)

29

Literatura

- Shortliffe, E. and Buchanan, B.: A model of inexact reasoning in medicine. Mathematical Biosciences, vol. 23, 351–379 (1975)

30

Dodatek. Współczynnik zależności: $P(X) + P(Y) \leq 1$

- **Tabela.** Porówanie wartości *lift*($X \rightarrow Y$), *cf*($X \rightarrow Y$) oraz *df*($X \rightarrow Y$), gdy $P(X) + P(Y) \leq 1$

P(*X*)	*P*(*Y*)	*P*(*XY*)	*P*(*X*)×*P*(*Y*)	*lift*(*X*→*Y*)	*cf*(*X*→*Y*)	*cf*(*Y*→*X*)	*df*(*X* →*Y*)=*df*(*Y*→*X*)
0.60	**0.30**	**0.30**	**0.18**	**1.67**	**0.29**	**1.00**	**1.00**
0.60	0.30	0.25	0.18	1.39	0.17	0.58	0.58
0.60	0.30	0.20	0.18	1.11	0.05	0.17	0.17
0.60	0.30	0.18	0.18	1.00	0.00	0.00	0.00
0.60	0.30	0.15	0.18	0.83	-0.17	-0.17	-0.17
0.60	0.30	0.10	0.18	0.56	-0.44	-0.44	-0.44
0.60	0.30	0.00	0.18	0.00	-1.00	-1.00	-1.00

Dodatek. Współczynnik zależności: $P(X) + P(Y) > 1$

- **Tabela.** Porówanie wartości *lift*($X \rightarrow Y$), *cf*($X \rightarrow Y$) oraz *df*($X \rightarrow Y$), gdy $P(X) + P(Y) > 1$

P(*X*)	*P*(*Y*)	*P*(*XY*)	*P*(*X*)×*P*(*Y*)	*lift*(*X*→*Y*)	*cf*(*X*→*Y*)	*cf*(*Y*→*X*)	*df*(*X* →*Y*)=*df*(*Y*→*X*)
0.80	0.60	0.60	0.48	1.25	0.38	1.00	1.00
0.80	0.60	0.55	0.48	1.15	0.22	0.58	0.58
0.80	0.60	0.50	0.48	1.04	0.06	0.17	0.17
0.80	0.60	0.48	0.48	1.00	0.00	0.00	0.00
0.80	0.60	0.45	0.48	0.94	-0.06	-0.06	-0.37
0.80	**0.60**	**0.40**	**0.48**	**0.83**	**-0.17**	**-0.17**	**-1.00**