## Interestingness/Evaluation Measures of Association Rules

Marzena Kryszkiewicz Institute of Computer Science Warsaw University of Technology

## Relative support of an Association Rule

♦ The *relative support* of an association rule  $X \rightarrow Y$  is defined as the probability of the co-occurrence of X and Y:

$$rSup(X \to Y) = P(XY).$$

 Property. The relative support depends on the probability of the base of a rule, but depends neither on the probability of its antecedent nor the probability of its consequent.

2

## Confidence of an Association Rule

◆ The confidence of an association rule X → Y is defined as the conditional probability that Y occurs provided X occurs:

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

 Property. The confidence depends on the probability of the antecedent of a rule and the probability of its base, but does not depend on the probability of its consequent.

## (In)dependence of Events

◆ *X* and *Y* called *independent* if:

$$P(XY) = P(X) * P(Y).$$

- Otherwise, they are called dependent.
- **♦** Natural interpretation:
  - If P(XY) > P(X) \* P(Y), then X and Y are dependent positively.
  - If P(XY) = P(X) \* P(Y), then X and Y are independent.
  - If P(XY) < P(X) \* P(Y), then X and Y are dependent negatively.
- ◆ **Property.**  $P(XY) \neq P(X) * P(Y)$  does not determine the degree to which *X* and *Y* are dependent.

## Confidence and event (in)dependence...

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

- ♦  $P(X) = \frac{2}{10}$ ,  $P(Y) = \frac{9}{10} \Rightarrow P(X) \times P(Y) = 0.18$ .
- $P(XY) = \frac{1}{10} = 0.1 < 0.18 = P(X) \times P(Y)$ , czyli X and Y are dependent negatively.
- $\bullet \ conf(X \to Y) = P(XY) / P(X) = \frac{1}{2}.$
- $P(X) = \frac{2}{10}, P(\overline{Y}) = \frac{1}{10} \Rightarrow P(X) \times P(\overline{Y}) = 0.02.$
- ♦  $P(X\overline{Y}) = \frac{1}{10} = 0.1 > 0.02 = P(X) \times P(\overline{Y})$ , czyli X and  $\overline{Y}$  are dependent positively.
- $\bullet \ conf(X \to \overline{Y}) = P(X\overline{Y}) \ / \ P(X) = \frac{1}{2}.$

## Confidence and event (in)dependence

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 and Y are dependent negatively.
- $\bullet \ conf(X \to Y) = P(XY) / P(X) = \frac{2}{3}.$
- $P(X) = \frac{3}{10}, P(\overline{Y}) = \frac{1}{10} \Rightarrow P(X) \times P(\overline{Y}) = 0.03.$
- ♦  $P(X\overline{Y}) = \frac{1}{10} = 0.1 > 0.03 = P(X) \times P(\overline{Y})$ , czyli X and  $\overline{Y}$  are dependent positively.
- $\bullet \ conf(X \to \overline{Y}) = P(X\overline{Y}) \ / \ P(X) = \frac{1}{3}.$

### Lift of an Association Rule

◆ The *lift* of an association rule X → Y is defined as the ratio of the conditional probability of the occurrence of Y given X occurs to the probability of the occurrence of Y:

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

• Property.

Y and X are dependent pos.	$P(XY) > P(X) \times P(Y)$	$lift(X \rightarrow Y) > 1$	$lift(Y \rightarrow X) > 1$
Y and X are independent	$P(XY) = P(X) \times P(Y)$	$lift(X \rightarrow Y) = 1$	$lift(Y \rightarrow X) = 1$
Y and X are dependent neg.	$P(XY) < P(X) \times P(Y)$	$lift(X \rightarrow Y) < 1$	$lift(Y \rightarrow X) < 1$

• **Property.**  $lift(X \rightarrow Y) = lift(Y \rightarrow X)$ .

## Lift and event (in)dependence | 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 | | 9 | x | x | | 9 | x | x | | 10 | X | Y | Z | V | | • $P(X) = \frac{9}{10}, P(Y) = \frac{9}{10} \Rightarrow P(X) \times P(Y) = 0.81.$ • • $P(XY) = \frac{9}{10} = 0.9 > 0.81 = P(X) \times P(Y), \text{ czyli}$ | • $P(XY) = \frac{9}{10} = 0.9 > 0.81 = P(X) \times P(Y), \text{ czyli}$ | • $P(XY) = \frac{9}{10} = 0.9 > 0.81 = P(X) \times P(Y), \text{ czyli}$ | • $P(XY) = \frac{1}{10} = 0.1 > 0.01 = P(X) \times P(Y) = 0.01.$ • • $P(XY) = \frac{1}{10} = 0.1 > 0.01 = P(X) \times P(Y), \text{ czyli}$ | • $P(XY) = \frac{1}{10} = 0.1 > 0.01 = P(X) \times P(Y), \text{ czyli}$ | • $P(XY) = \frac{1}{10} = 0.1 > 0.01 = P(X) \times P(Y), \text{ czyli}$ | • $P(XY) = \frac{1}{10} = 0.1 > 0.01 = P(X) \times P(Y), \text{ czyli}$ | • $P(XY) = \frac{1}{10} = 0.1 > 0.01 = P(X) \times P(Y), \text{ czyli}$ | • $P(XY) = \frac{1}{10} = 0.1 > 0.01 = P(X) \times P(Y), \text{ czyli}$ | • $P(XY) = \frac{1}{10} = 0.1 > 0.01 = P(X) \times P(Y), \text{ czyli}$ | • $P(XY) = \frac{1}{10} = 0.1 > 0.01 = P(X) \times P(Y), \text{ czyli}$ | • $P(XY) = \frac{1}{10} = 0.1 > 0.01 = P(X) \times P(Y), \text{ czyli}$ | • $P(XY) = \frac{1}{10} = 0.1 > 0.01 = P(X) \times P(Y), \text{ czyli}$ | • $P(XY) = \frac{1}{10} = 0.1 > 0.01 = P(X) \times P(Y), \text{ czyli}$

10 x x

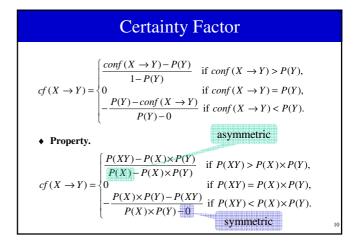
•  $lift(Z \to V) = \frac{P(ZV)}{P(Z) \times P(V)} = \frac{0.1}{0.1 \times 0.1} = \frac{10}{1} = 10.$ 

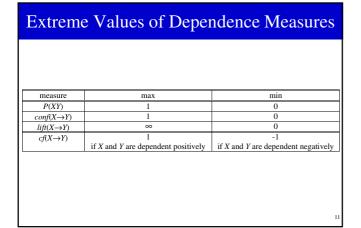
## Certainty Factor of an Association Rule...

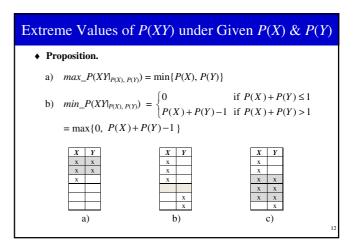
◆ The *certainty factor* of an association rule X → Y is defined as the degree to which the probability of the occurrence of Y can change when X occurs:

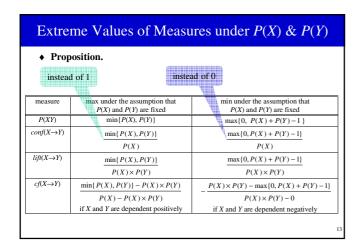
$$cf(X \to Y) = \begin{cases} \frac{conf(X \to Y) - P(Y)}{1 - P(Y)} & \text{if } conf(X \to Y) > P(Y), \\ 0 & \text{if } conf(X \to Y) = P(Y), \\ -\frac{P(Y) - conf(X \to Y)}{P(Y) - 0} & \text{if } conf(X \to Y) < P(Y). \end{cases}$$

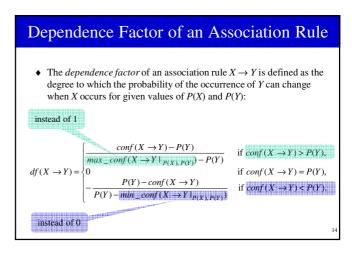


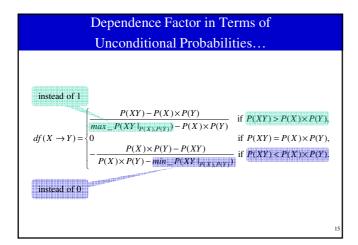


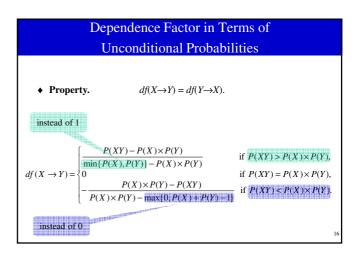












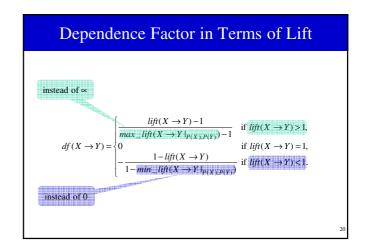
# Dependence Factor versus Certainty Factor ◆ Theorem. > df(X o Y) = cf(X o Y) = 0 if $P(XY) = P(X) \times P(Y)$ . > $|df(X o Y)| \ge |cf(X o Y)|$ if $P(XY) \ne P(X) \times P(Y)$ . > If $P(XY) > P(X) \times P(Y)$ , then: $df(X o Y) = \max\{cf(X o Y), cf(Y o X)\}.$ > If $P(XY) < P(X) \times P(Y)$ , then: df(X o Y) = cf(X o Y) if $P(X) + P(Y) \le 1$ . |df(X o Y)| > |cf(X o Y)| if P(X) + P(Y) > 1.

Dependence Factor: $P(X) + P(Y) < 1$ • Table. Comparison of values of $lift(X \rightarrow Y)$ , $cf(X \rightarrow Y)$ and $df(X \rightarrow Y)$							
who	P(Y) - 0.30	$+P(Y) < \frac{P(XY)}{0.30}$	P(X)×P(Y) 0.18	lift(X→Y) <b>1.67</b>	<i>cf(X→Y)</i> <b>0.29</b>	cf(Y→X)	$\frac{df(X \to Y) = df(Y \to X)}{1.00}$
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

## Dependence Factor: P(X) + P(Y) > 1

◆ **Table.** Comparison of values of  $lift(X \rightarrow Y)$ ,  $cf(X \rightarrow Y)$  and  $df(X \rightarrow Y)$  when P(X) + P(Y) > 1

P(X)	P(Y)	P(XY)	$P(X)\times P(Y)$	$lift(X \rightarrow Y)$	$cf(X \rightarrow Y)$	$cf(Y \rightarrow X)$	$df(X \rightarrow Y) = df(Y \rightarrow X)$
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



## Conclusions

- ◆ df(X→Y) always reaches 1 when the dependence between X and Y is strongest possible positive and -1 when the dependence between X and Y is strongest possible negative for any given values of P(X) and P(Y).
- Unlike the dependence factor, the certainty factor and lift are misleading in expressing the strength of the dependence as each of them may take a value close to the value characteristic for independence even when two events are maximally strongly positively or negatively dependent.

## Further results

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

21

## References

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

23