

# Itemset & Rule Assessment

Let  $I$  be a set of items

$T$  be a set txn ids

$D \subseteq T \times I$  be a binary DB

$F$  is a frequent itemset

Recall rule  $X \rightarrow Y$

$\xrightarrow{\text{consequent}}$

$\xleftarrow{\text{antecedent}}$

$$t(X) = \{t \in T \mid X \text{ is contained in the txn}\}$$

Rule assessment

support  $\sup(X \rightarrow Y) = \sup(XY) = |t(XY)|$

rel support  $r\sup(X \rightarrow Y) = \frac{\sup(XY)}{|D|} = P(XY)$

confidence  $\text{conf}(X \rightarrow Y) = \frac{r\sup(XY)}{r\sup(X)} = \frac{\sup(XY)}{\sup(X)}$

$\xleftarrow{\# \text{ of txns}}$

||

$$\frac{P(XY)}{P(X)} = P(Y|X)$$

$$\text{conf}(X \rightarrow Y) \stackrel{?}{=} \text{conf}(Y \rightarrow X)$$

is  $r\sup(X \rightarrow Y)$  sym? Yes

$$r\sup(X \rightarrow Y) = \frac{\sup(XY)}{|D|} = \frac{\sup(YX)}{|D|}$$

" "

$$r\sup(Y \rightarrow X)$$

Lift "surprise of a rule"

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

interpretation

$\sim 1 \Rightarrow$  expected

$P(XY) = P(X)P(Y)$  when  $X \& Y$  are ind

$\gg 1 \Rightarrow$  above expectation

$\ll 1 \Rightarrow$  below expectation

Props!

1.  $\text{lift}(X \rightarrow Y) = \text{lift}(Y \rightarrow X)$

2.  $\text{lift}(X \rightarrow Y) \geq \text{conf}(X \rightarrow Y)$

3. Not "downward closed"

consider  $\text{lift}(X \rightarrow Y)$

may  $\exists X' \subset X \& Y' \subset Y$

s.t.  $\text{lift}(X' \rightarrow Y') > \text{lift}(X \rightarrow Y)$

Warning!

Lift can be noisy in small datasets

Rule		lift
AE	$\rightarrow$	BC
CE	$\rightarrow$	AB
BE	$\rightarrow$	AC



Leverage : diff between observed & expected joint prob of XY assuming X & Y ind

$$\text{leverage}(X \rightarrow Y) = P(XY) - P(X) \cdot P(Y) = \text{rsup}(XY) - \text{rsup}(X) \cdot \text{rsup}(Y)$$

Using together

	Rule	rsup	lift	leverage
1.	$ACD \rightarrow E$	0.17	1.20	0.03
2.	$AC \rightarrow E$	0.33	1.20	0.06
3.	$\cancel{AB} \rightarrow D$	0.50	1.12	0.06
4.	$A \rightarrow E$	0.67	1.20	0.11

## Itemset Assessment Measures

Given itemset  $X = \{x_1, x_2, \dots, x_k\}$   
and binary DB  $D$

support  $\text{sup}(X) = |t(x)|$

relative support:  $\text{rsup}(X) = \frac{\text{sup}(X)}{|D|} = P(X)$

Lift:  $\text{lift}(X) = \frac{P(X)}{\prod_{i=1}^k P(x_i)} = \frac{\text{rsup}(X)}{\prod_{i=1}^k \text{rsup}(x_i)}$

## Rule-based Measure

Let  $\Theta$  be a rule assessment measure (conf, lift, leverage, etc...)

1. Generate all rules  $X_1 \rightarrow X_2$  &  $X_2 \rightarrow X_1$  in which  
 $X_1, X_2 \subseteq X$  and  $X_1 \cap X_2 = \emptyset$

(define  $\subseteq$  as all bipartitions of  $X$ )

etc

2. take  $\text{avg}$ ,  $\text{max}$ ,  $\min$  over  $\Theta$

$$\text{Avg } \Theta = \text{avg}_{X_1, X_2 \subseteq X} \left\{ \Theta(X_1 \rightarrow X_2) \right\}$$

concrete example  $\text{MaxLift}(X)$   
 $= \max_{X_1, X_2 \subseteq X} \left\{ \text{lift}(X_1 \rightarrow X_2) \right\}$

$$\text{Max } \Theta = \max_{X_1, X_2 \subseteq X} \left\{ \Theta(X_1 \rightarrow X_2) \right\}$$