

Association Analysis

Exercise 6



Frequent Itemsets & Association Rules

- Frequent Itemset
 - Support count: Frequency of an itemset
 - Support: relative frequency of an itemset (wrt. all transactions)
- Association Rule $X \rightarrow Y$
 - Support: Support of the itemset $X \cup Y$
 - Confidence: relative frequency of $X \cup Y$ wrt. X
 - “If an itemset contains X , in x% of the cases it also contains Y ”
 - Lift: confidence of rule $X \rightarrow Y$ divided by support of consequent Y
 - >1 X and Y are positively correlated
 - <1 X and Y are negatively correlated
 - $=1$ X and Y are independent

The Apriori Principle


- If an itemset is frequent, then all of its subsets must also be frequent
 - Support of an itemset never exceeds the support of its subsets
 - “anti-monotone” property of support
- Example:
 - 50% of all shopping baskets contain beer
 - Then beer & nachos can’t be in more than 50% of all shopping baskets
 - So if any itemset is less frequent than we require, no need to check larger itemsets

$$\forall X, Y : (X \subseteq Y) \Rightarrow s(X) \geq s(Y)$$

Operators: FP-Growth

- Input Port
 - Example Set (binominal attributes!)
- Output Ports
 - Example Set
 - Frequent Item Sets
- Parameters
 - If “find min number of itemsets”
 - Min number of itemsets
 - Else
 - Min support
 - Positive value
 - Max items
 - Must contain

Parameters ×

 **FP-Growth**

☒ find min number of itemsets ✓ ⓘ

min number of itemsets

100 ⓘ

max number of retries

15 ⓘ

positive value

ⓘ

min support

0.95 ⓘ

max items ✓

-1 ⓘ

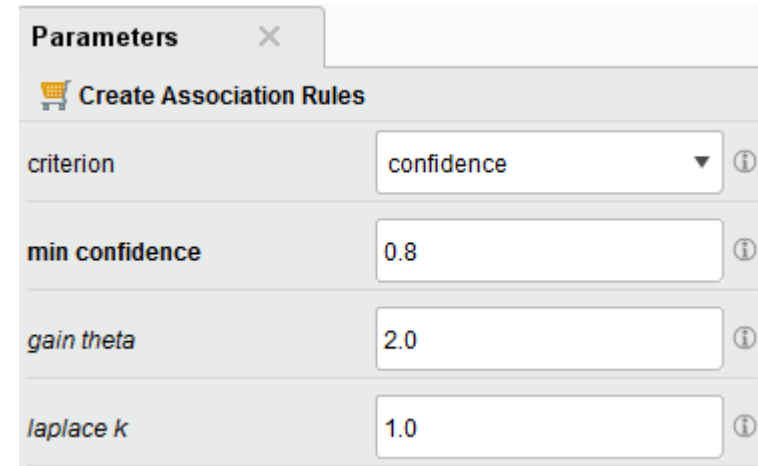
must contain

ⓘ



Operators: Create Association Rules

- Input Port
 - Frequent Item Sets
- Output Ports
 - Association Rules
 - Frequent Item Sets
- Parameters
 - Criterion
 - Min criterion value
 - Gain theta (used if criterion = gain)
 - Laplace k (used if criterion = laplace)



The screenshot shows a software interface for configuring the 'Create Association Rules' operator. It features a title bar with a close button and a shopping cart icon. Below the title bar, there are four parameter rows, each with a label, a value field, and an information icon. The parameters are: 'criterion' set to 'confidence', 'min confidence' set to '0.8', 'gain theta' set to '2.0', and 'laplace k' set to '1.0'.

Parameter	Value
criterion	confidence
min confidence	0.8
gain theta	2.0
laplace k	1.0

