Association rules

Association rules are rules of the form **X → Y**, where X and Y are conjunctions of items. Task: Find **all** association rules that satisfy minimum support and minimum confidence constraints.

Support: *Sup (X → Y) = #XY / #D ≅ p(XY)*

Confidence: *Conf (X → Y) = #XY / #X ≅ p(XY) / p(X) = p(Y|X)*

The **anti-monotone property of support:** if we drop out an item from an itemset, support value of new itemset generated will either be the same or will increase.

$$\forall A, B : A \subseteq B \Rightarrow supp(A) \geq supp(B)$$

In general, confidence does not have an anti-monotone property: Conf(ABC → D) can be larger or smaller than Conf(AB → D). Confidence of rules generated from the same itemset has an anti-monotone property:

Conf ({A,B} → {C}) >= Conf ({A} → {B,C})

Conf({A,B,C} → {D}) ≥ Conf({A,B} → {C,D}) ≥ Conf({A} → {B,C,D})

Association rules algorithm Apriori (data, minSupport, minConfidence)

1. Generate frequent itemsets with a minimum support constraint.
2. Generate rules from frequent itemsets with a minimum confidence constraint.

Frequent itemsets mining

Items in an itemset should **always** be sorted alphabetically.

- Generate all 1-itemsets with the given minimum support.
- Use 1-itemsets to generate 2-itemsets with the given minimum support.
- From n-itemsets generate (n+1)-itemsets as unions of n-itemsets with the same (n-1) items at the beginning & prune* to satisfy the minimum support constraint.

*prune: Remove an itemset if some of its subsets are not present at the previous level, remove an itemset if the support count in the datset does not satisfy the minimum support constraint.

The result is a lattice of frequent itemsets.

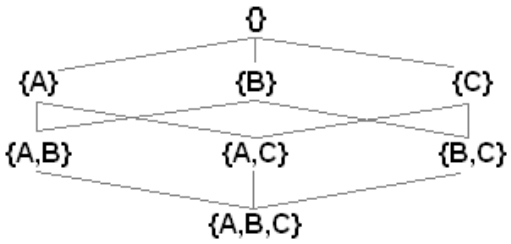


Figure 1: A lattice of items A, B and C

Petra Kralj Novak
Petra.Kralj.Novak@ijs.si
http://kt.ijs.si/petra_kralj/dmkd.html

Generating rules from frequent itemsets

The support of the rule is the same as the support of the itemset it was constructed from. From each frequent itemset construct all possible rules with at least one item on the left and one on the right by taking into account that transferring members of a supported itemset from the left-hand side of a rule to the right-hand side cannot increase the value of rule confidence. All the counts are in the lattice.

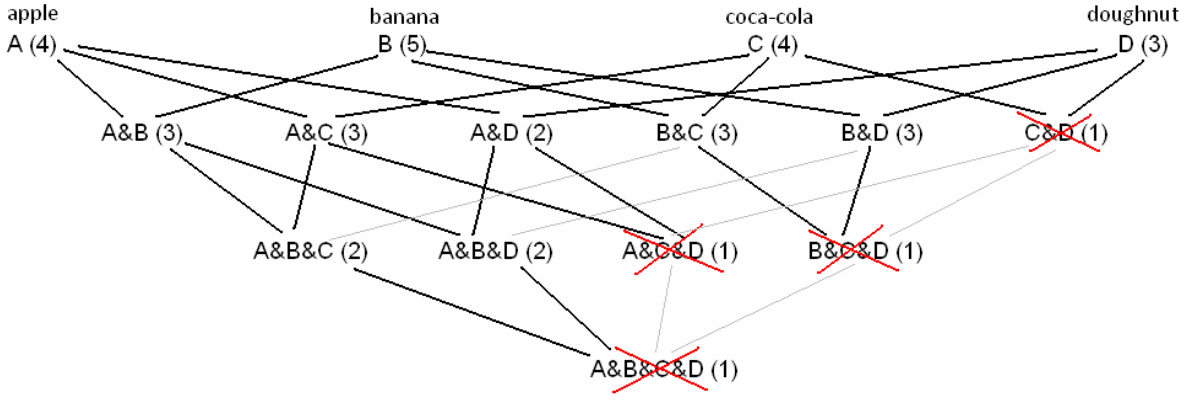
Exercise

Find all association rules with minimum support 2/6 and minimum confidence 70%.

A	B	C	D
1	1	1	1
	1	1	
	1		1
1		1	
1	1		1
1	1	1	

A = apple
B = banana
C = coca-cola
D = doughnut

The itemset lattice:



Note:

1. In the itemset lattice, the support of the itemsets decreases along the path from top to bottom of the lattice.
2. The lattice gives a partial order of the itemsets according to the subsumption operator.
3. Since items are sorted in every itemset, when generating 3-itemsets from 2-itemsets, it is enough to merge the 2-itemsets that have the same first item. Similarly, when generating 4-itemsets from 3-itemsets, we couple 3-itemsets with the same first two items, and so on. This guarantees that the same itemset is not generated twice.
4. If a subset on an itemset does not appear on the previous level, it is not supported.
5. If an itemset does not fulfill the minimum support constraint, it is discarded and not used when constructing itemsets at the next level.

Petra Kralj Novak
Petra.Kralj.Novak@ijs.si
http://kt.ijs.si/petra_kralj/dmkd.html

Generating rules from frequent itemsets:

Itemset (count)	Rule	Support	Confidence	Over threshold
AB (3)	A → B	3/6	3/4 = 75%	✔
	B → A	3/6	3/5 = 60%	
AC (3)	A → C	3/6	3/4 = 75%	✔
	C → A	3/6	3/4 = 75%	✔
AD (2)	A → D	2/6	2/4 = 50%	
	D → A	2/6	2/3 = 67%	
BC (3)	B → C	3/6	3/5 = 60%	
	C → B	3/6	3/4 = 75%	✔
BD (3)	B → D	3/6	3/5 = 60%	
	D → B	3/6	3/3 = 100%	✔
ABC (2)	AB → C	2/6	2/3 = 67%	
	AC → B	2/6	2/3 = 67%	
	BC → A	2/6	2/3 = 67%	
	A → BC	We do not generate these rules because transferring members of a supported itemset from the left-hand side of a rule to the right-hand side cannot increase the value of rule confidence.		
	B → AC			
	C → AB			
ABD (2)	AB → D	2/6	2/3 = 67%	
	AD → B	2/6	2/2 = 100%	✔
	BD → A	2/6	2/3 = 67%	
	A → BD	2/6	2/4 = 50%	
	B → AD	We do not generate this rule.		
	D → AB	2/6	2/3 = 67%	

Note:

1. All the rules deriving from the same itemset have the same support
2. All the counts (supports) for computing the confidence of rules are in the itemset lattice.

Petra Kralj Novak
Petra.Kralj.Novak@ijs.si
http://kt.ijs.si/petra_kralj/dmkd.html