

Week 2 and 3: Tutorial exercises (March 17th and 21st 2006) Association Rule Mining.

Exercise 1. Apriori

Trace the results of using the Apriori algorithm on the grocery store example with support threshold $s=33.34\%$ and confidence threshold $c=60\%$. Show the candidate and frequent itemsets for each database scan. Enumerate all the final frequent itemsets. Also indicate the association rules that are generated and highlight the strong ones, sort them by confidence.

Transaction ID	Items
T1	HotDogs, Buns, Ketchup
T2	HotDogs, Buns
T3	HotDogs, Coke, Chips
T4	Chips, Coke
T5	Chips, Ketchup
T6	HotDogs, Coke, Chips

Exercise 2. FP-tree and FP-Growth

- a) Use the transactional database from the previous exercise with same support threshold and build a frequent pattern tree (FP-Tree). Show for each transaction how the tree evolves.
- b) Use Fp-Growth to discover the frequent itemsets from this FP-tree.

Exercise 3: Using WEKA

Load a dataset described with nominal attributes, e.g. weather.nominal. Run the Apriori algorithm to generate association rules.

Exercise 4: Apriori and FP-Growth (to be done at your own time, not in class)

Giving the following database with 5 transactions and a minimum support threshold of 60% and a minimum confidence threshold of 80%, find all frequent itemsets using (a) Apriori and (b) FP-Growth. (c) Compare the efficiency of both processes. (d) List all strong association rules that contain "A" in the antecedent (Constraint). (e) Can we use this constraint in the frequent itemset generation phase?

TID	Transaction
T1	{A, B, C, D, E, F}
T2	{B, C, D, E, F, G}
T3	{A, D, E, H}
T4	{A, D, F, I, J}
T5	{B, D, E, K}