

Data Mining Homework - Association Analysis

Instructions

Please finish the homework assigned below and submit your homework electronically. Any submissions after the deadline will be given a zero grade.

Questions

1. In the following, we have ten market-basket transactions:

Transaction ID	: Items Bought
1	: {A,B,D,E}
2	: {B,C,D}
3	: {A,B,D,E}
4	: {A,C,D,E}
5	: {B,C,D,E}
6	: {B,D,E}
7	: {C,D}
8	: {A,B,C}
9	: {A,D,E}
10	: {B,D}

- (a) What is the maximum size of frequent itemsets that can be extracted (assuming $minsup = 0$)?
 - (b) What is the maximum number of association rules that can be extracted from this data set (including rules that have zero support)?
 - (c) Compute the support, confidence, and lift for the association rule $\{A, D\} \rightarrow \{E\}$
 - (d) Find all the frequent itemsets assuming $minsup\ count = 2$.
 - (e) Find an itemset (of size 2 or larger) that has the largest support.
 - (f) Find a pair of items, say x and y , such that $\{x\} \rightarrow \{y\}$ and $\{y\} \rightarrow \{x\}$ have the same confidence.
2. Given a k -itemset X and an l -itemset $Y \subset X$ from a set of transactions T , show that:

If Y is infrequent and $l < k - 1$, then there exists an infrequent $(k - 1)$ -itemset $Z \subset X$.

Note that, the notation $A \subset B$ indicates ‘A is a subset of B’.

Hints: If Y is infrequent then any superset of Y will also be infrequent.

Your proof can be just intuitive explanation with or without mathematical deductions. You can also try the lattice graph we introduced in class.

Why is the claim important? It’s useful because we only need to check the $(k - 1)$ -subsets of the k -itemset X to decide whether we have to count its frequency: If there exists an infrequent $(k - 1)$ subset, we are sure that X is also infrequent so we should not count its frequency. If all $(k - 1)$ subsets are frequent, we have to count the frequency of X .