

Data Mining

Association Rule Mining (Part B)

Dr. Jason T.L. Wang, Professor
Department of Computer Science
New Jersey Institute of Technology

Where am I?

Association rule:

Banana → Orange (10%, 50%)

Goal: Find all association rules satisfying minimum support and minimum confidence requirements

Solution: Use Apriori algorithm to generate all frequent itemsets

Apriori Principle

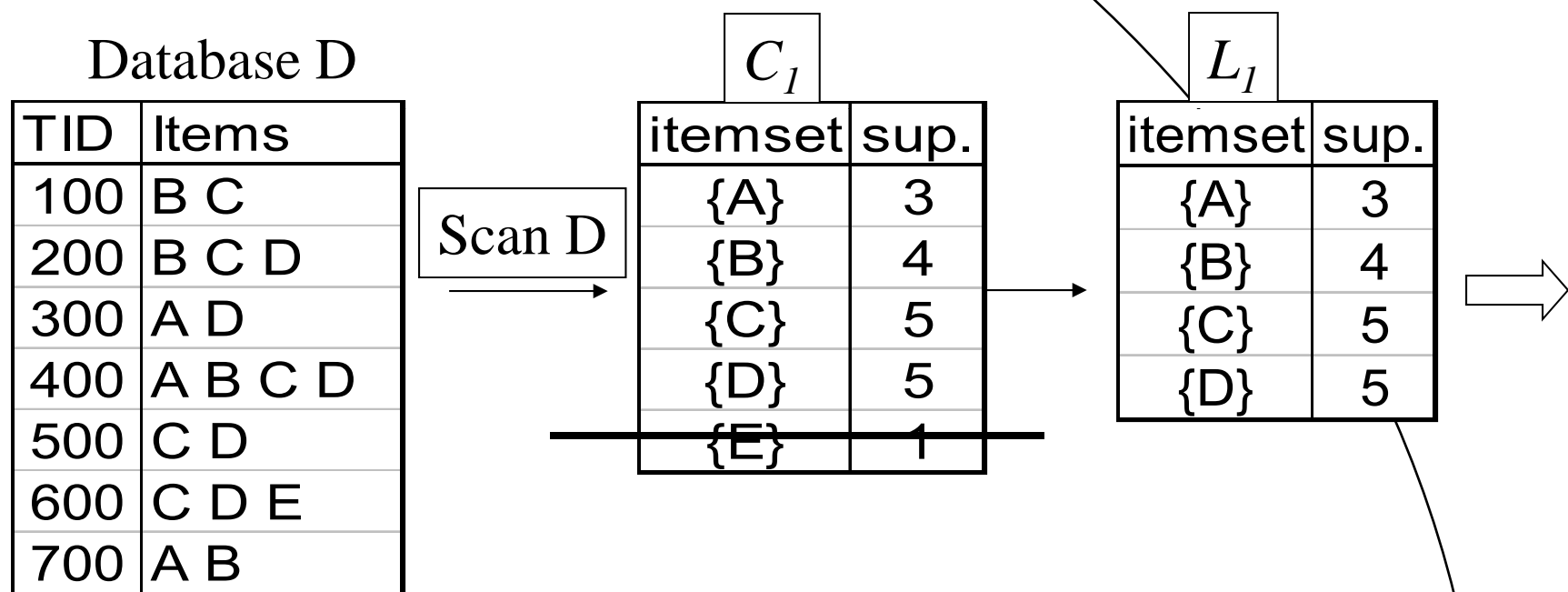
- Any subset of a frequent itemset must be frequent.
- Any superset of a non-frequent itemset must be non-frequent.

Apriori Algorithm

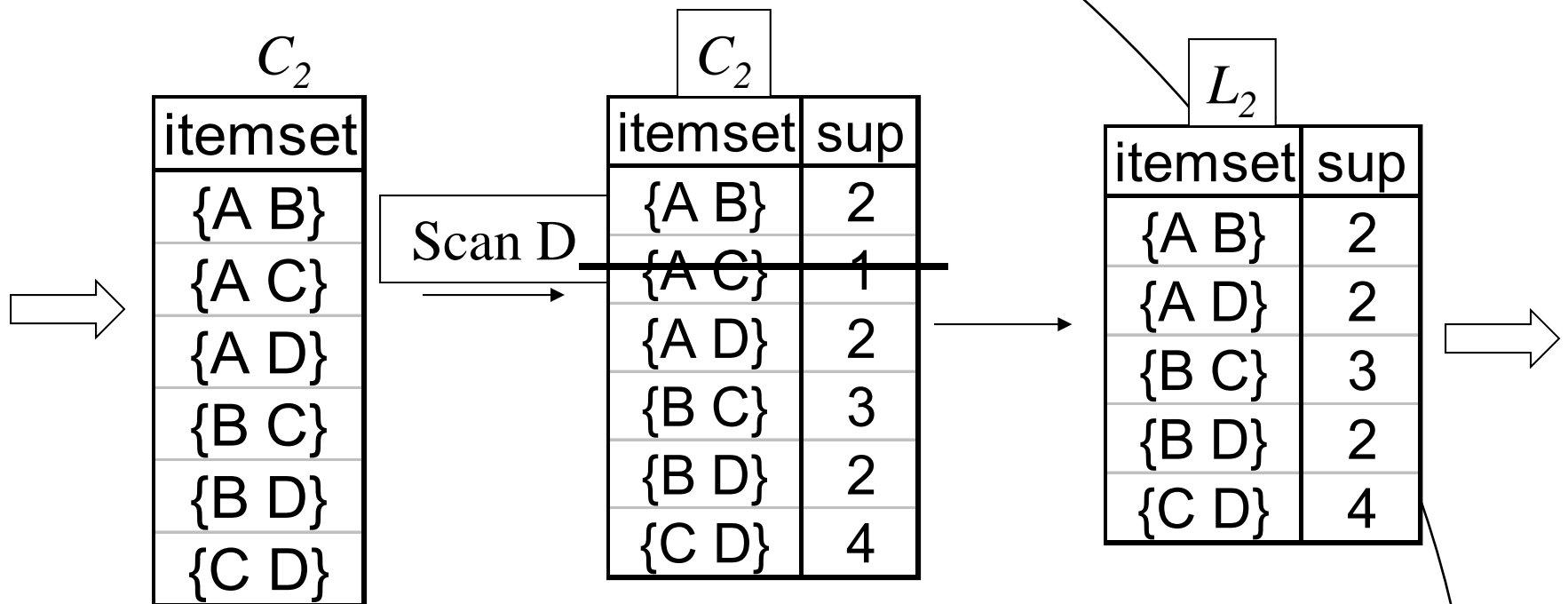
- Find the sets of items that have minimum support (frequent itemsets) starting from 1-itemsets and expanding to k-itemsets; if a j-itemset is already not frequent, then do not consider any superset of it.
- Use the frequent itemsets to generate all association rules.

The Apriori Algorithm Example

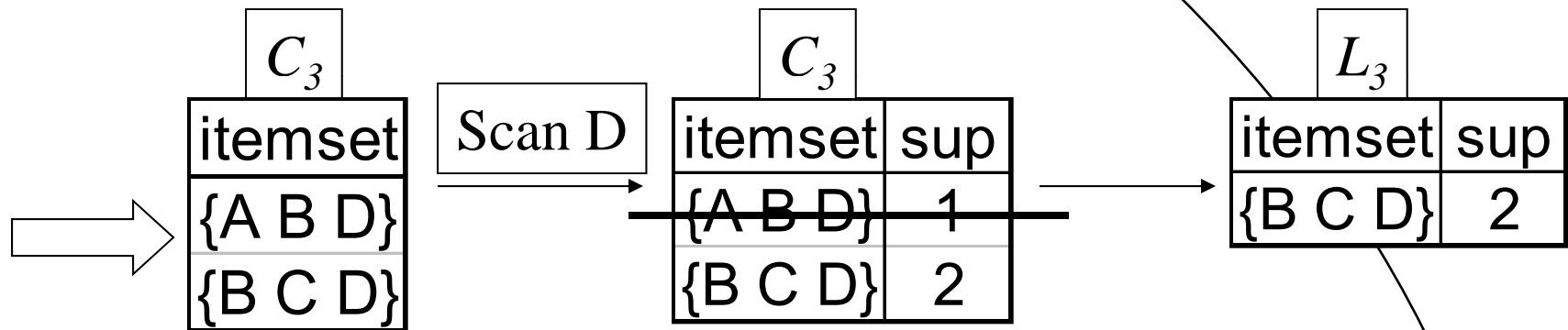
(minimum support $\geq 20\%$)



Example – cont'd



Example – cont'd



Example – cont'd

Minimum confidence $\geq 50\%$

Consider frequent itemsets $\{B\ C\ D\}$, $\{A\ B\}$, $\{A\ D\}$, $\{A\}$, $\{B\}$, $\{C\}$, ...

Here are some association rules (no rule is generated from frequent 1-itemsets):

$B\ C \rightarrow D$ [28.6%, 66.7%]

$B\ D \rightarrow C$ [28.6%, 100%]

$A \rightarrow B$ [28.6%, 66.7%]

...



End of Association Rule Mining Module (Part B)