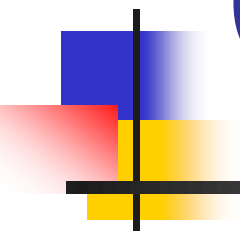


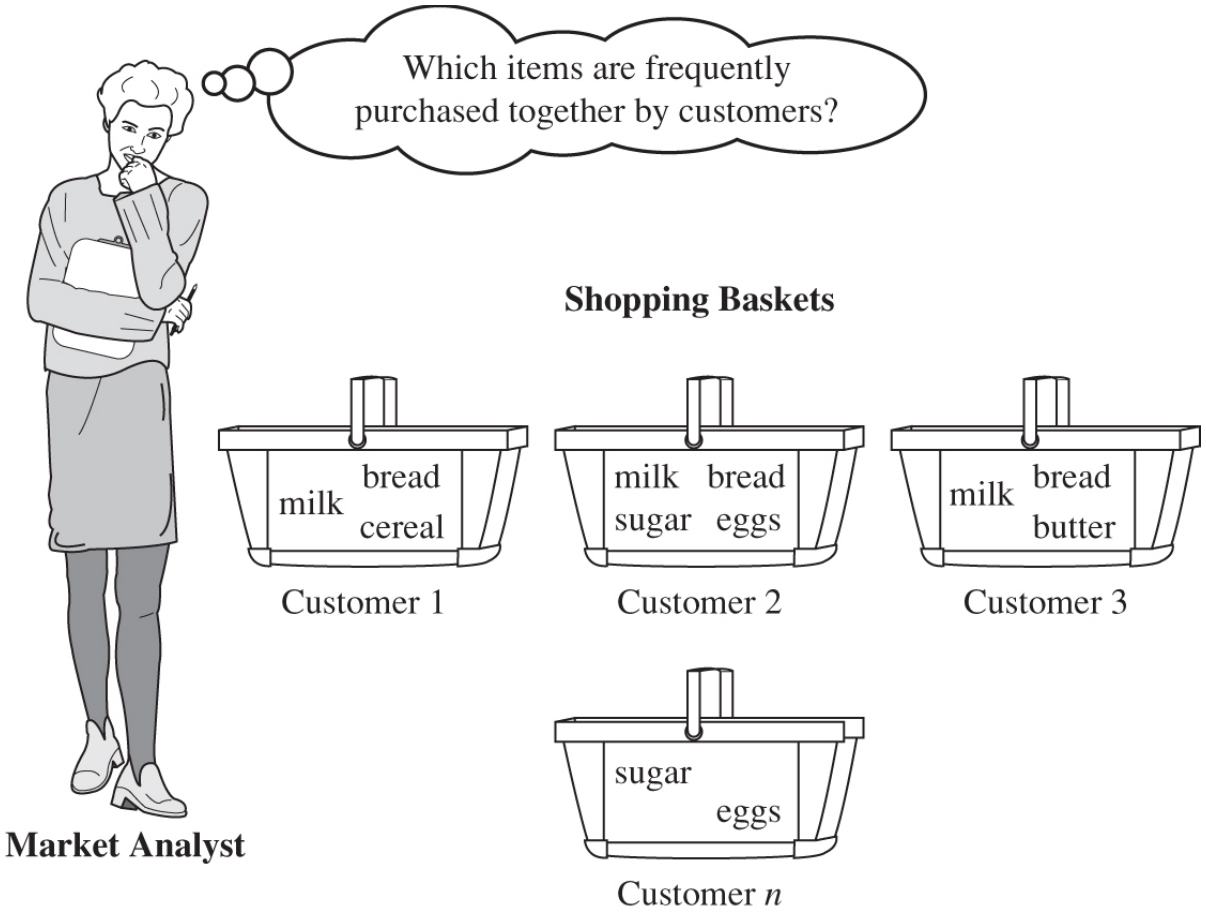
CS 654

Topic:
Association Rule Mining

Association Rule Mining

- Association Mining
- Shopping patterns discovery
- Basket/Shopping Cart Analysis
- Want to know what items been purchased together.
- Many applications for retails stores, e-commerce companies, …, etc.
  - Give a few examples yourself.

Shopping Patterns Discovery



Basic Concepts

- View each basket as one transaction.
- A transaction consists of items.
- An example:

Transaction-id	Items bought
10	A, B, C
20	A, C
30	A, D
40	B, E, F

What Is Association Mining?

- Association rule mining:
  - Finding frequent patterns, associations, correlations, or causal structures among sets of items or objects in transaction databases, relational databases, and other information repositories.
  - Frequent pattern: pattern (set of items) that occurs frequently in a database
  - What products were often purchased together?
  - We can use **Apriori Algorithm** to find the rules.

Basic Concepts (continued)

- X is an Itemset, i.e., contains one or more items.
- Y is an Itemset, i.e., contains one or more items.
- An association rule:

  X -> Y means if X was purchased, so was Y.
- Support (X->Y) = P(X U Y) (Both X and Y were purchased together in a transaction. Note: U means both.)
- Confidence (X->Y) = P(Y|X) = support(XUY)/support(X) = support_count(XUY)/support_count(X)
- Minimum support threshold
- Minimum confidence threshold
- Use 0% - 100% or 0 – 1.0 to denote the values.

Basic Concepts (continued)

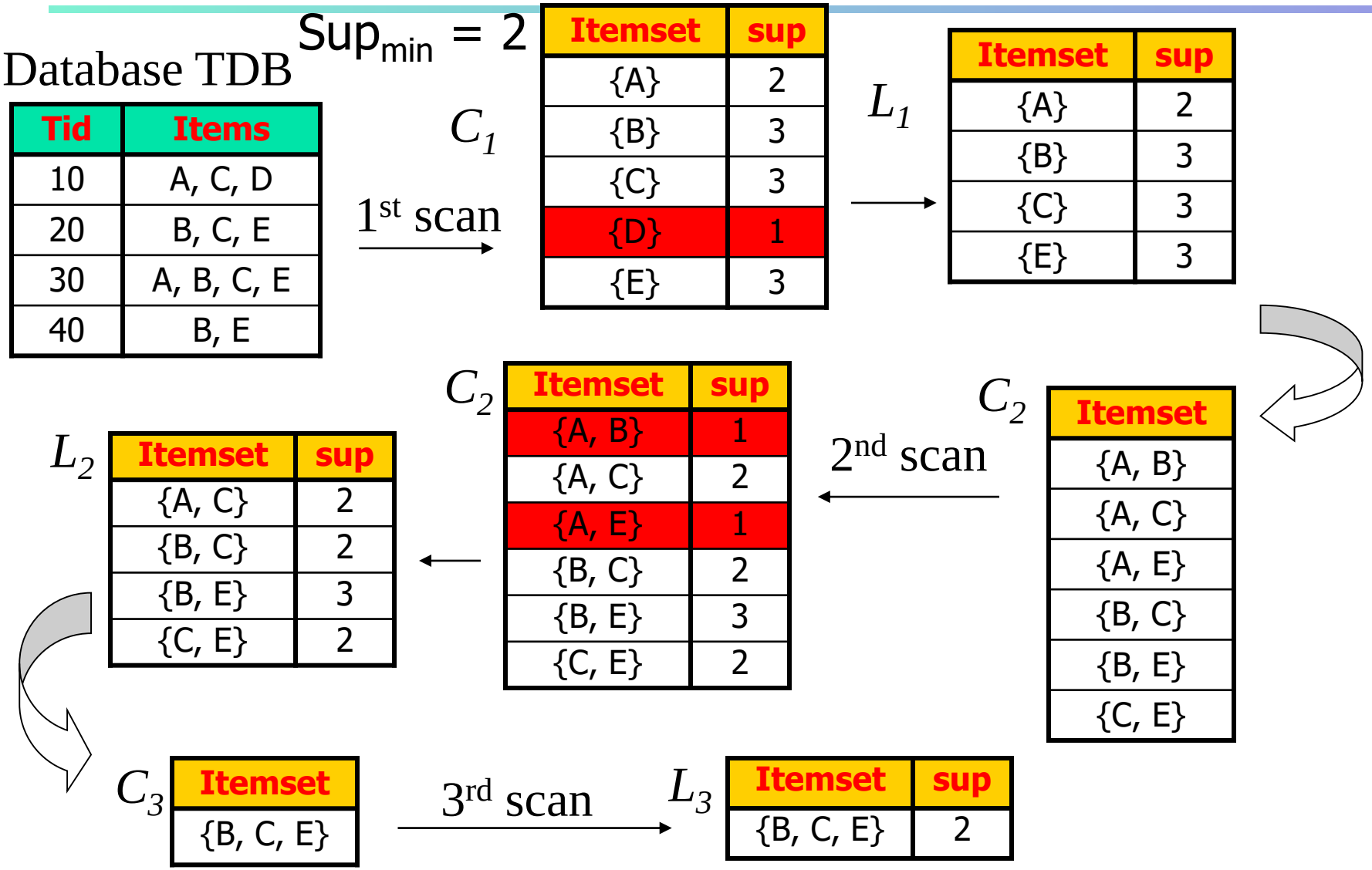
- Itemset $X=\{x_1, \ldots, x_k\}$
- Find all the rules $X \rightarrow Y$ with min confidence and support
  - support, *s*, probability that a transaction contains $X \cup Y$
  - confidence, *c*, conditional probability that a transaction having X also contains *Y*.
- We are interested in discovering rules that satisfy minimum support and confidence.

Illustrating example

Tid	Items
10	A, C, D
20	B, C, E
30	A, B, C, E
40	B, E

- There are 4 transactions in total.
- 5 items, i.e., A,B,C,D,E have been purchased.
- Some items been purchased together in one transaction.
- Some items been purchased in more than one transaction.
- We request min_support = 50% and min_conf = 50%
  (normally given by the users).

The Apriori Algorithm—An Example



Rule: A -> C

Min. support 50%
Min. confidence 50%

For rule $A \Rightarrow C$:

support = support($\{A\} \cup \{C\}$) = 50%

confidence = support($\{A\} \cup \{C\}$)/support($\{A\}$) = 100%

A->C is not necessarily equal to C->A

With min_support = 50% and min_conf = 50%.

$C \rightarrow A$

Support (C) = ¾ = 75%

Support($C \cup A$) =2/4 = 50%

Confidence = 0.5/0.75 = 67%

Practice

Tid	Items
10	A, C, D
20	B, C, E
30	A, B, C, E
40	B, E

Can you list more rules in this example?

Another Example

TID	List of items
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

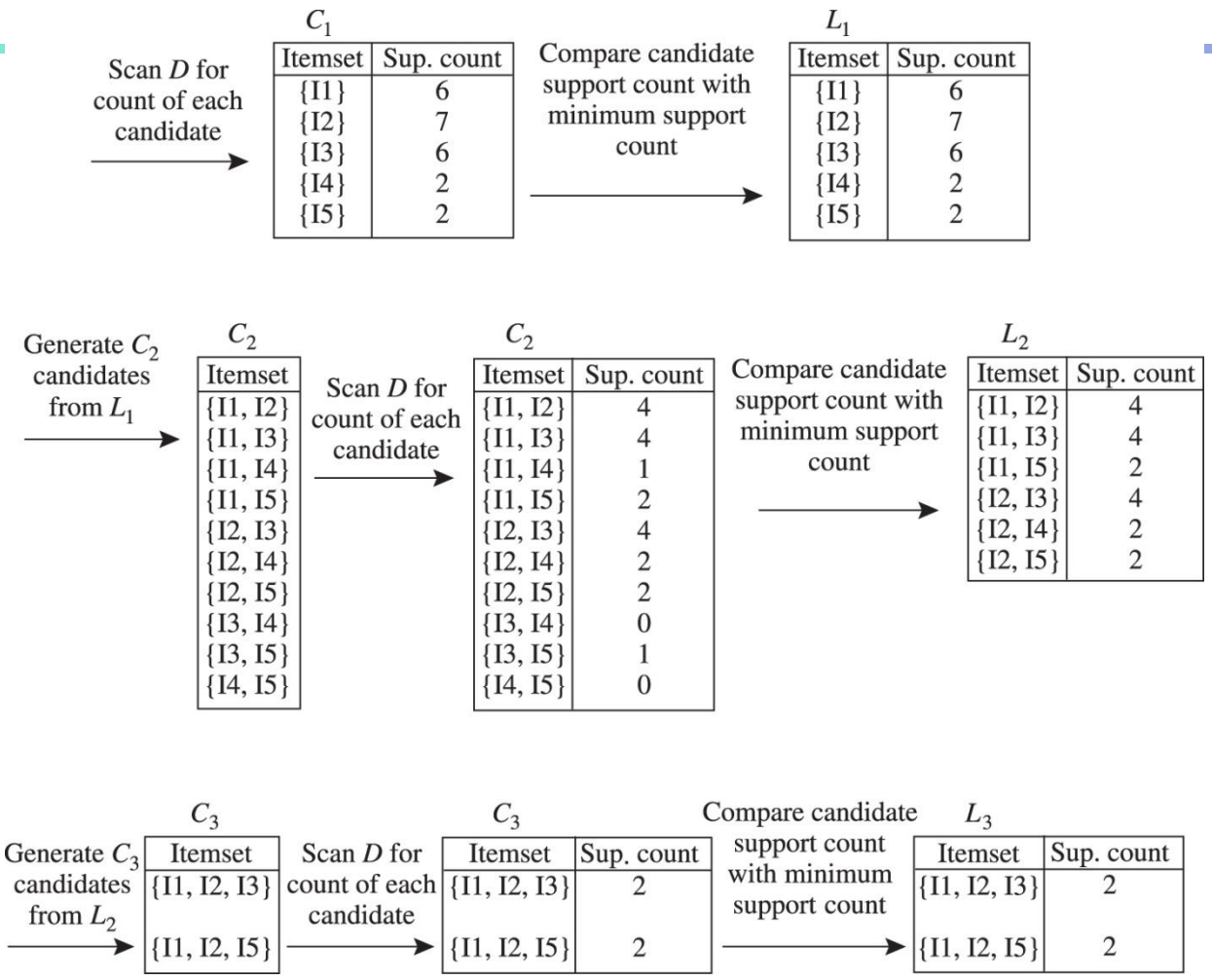


Figure 6.2 Generation of the candidate itemsets and frequent itemsets, where the minimum support count is 2.

Association Rules Generation

- Example:
  - Frequent itemset is {I1, I2, I5}

    Rule 1: {I1, I2} -> I5, confidence= 2/4

    Rule 2: {I1, I5} -> I2, confidence = 2/2

    Rules 3, 4, 5, 6?

The Apriori Algorithm

- Pseudo-code:

  C_k: Candidate itemset of size k
  L_k : frequent itemset of size k

  L_1 = {frequent items};
  for (k = 1; L_k !=∅; k++) **do begin**
  C_{k+1} = candidates generated from L_k;
  for each transaction t in database do
  increment the count of all candidates in C_{k+1} that are contained in t
  L_{k+1} = candidates in C_{k+1} with min_support
  end
  return $\cup_k L_k$;