



# BASIC CONCEPTS IN AR MINING

**SYRACUSE UNIVERSITY**  
School of Information Studies

# FREQUENT ITEMSET

Transaction ID	Items Bought *
10	A, B, D
20	A, C, D
30	A, D, E
40	B, E, F
50	B, C, D, E, F

Can you answer the following questions?

Which two items are frequently bought together?

Which three items are often bought together?

...

# DEFINITION: FREQUENT ITEMSET

Itemset:

A collection of one or more items

k-itemset contains k items

1-itemset:

{A}:3, {B}:3, {C}:2, {D}:4, {E}:3, {F}:2

2-itemset:

{A,B}:1, {A,D}:3

3-itemset:

{A,B,C}:0, {B,E,F}:2

Transaction ID	Items Bought
10	A, B, D
20	A, C, D
30	A, D, E
40	B, E, F
50	B, C, D, E, F

Frequently Bought Together



- ✓ **This item:** The Manga Guide to Database
- ✓ The Manga Guide to Statistics by Shin Taka
- ✓ The Manga Guide to Linear Algebra by Shi

# METRICS TO EVALUATE FREQUENT LEVEL OF ITEMSETS

How frequent is an itemset?

Support count:

Number of transactions that contain an itemset

$$\text{support\_count}(\{D, E\}) = 2$$

Support percentage:

Fraction of transactions that contain an itemset

$$\text{support}(\{D, E\}) = 2/5$$

Frequent itemset:

An itemset with  $\text{support} \geq \text{threshold}$

# DEFINITION: ASSOCIATION RULE

Association rule:

An implication of the form  $X \rightarrow Y$ ,  
where  $X$  and  $Y$  are itemsets,  
e.g.,  $\{E, F\} \rightarrow \{B\}$

Example Rules:

**LHS:**  
Left-  
Hand  
Side

→

$\{B, E\} \rightarrow \{F\}$

$\{E, F\} \rightarrow \{B\}$

$\{B, F\} \rightarrow \{E\}$

$\{B\} \rightarrow \{E, F\}$

$\{E\} \rightarrow \{B, F\}$

$\{F\} \rightarrow \{B, E\}$

**RHS:**

← Right-

Hand

Side

Transaction ID	Items Bought
10	A, B, D
20	A, C, D
30	A, D, E
40	B, E, F
50	B, C, D, E, F

# METRICS TO EVALUATE THE RULE'S STRENGTH

## Rule evaluation metrics

### Support $P(X, Y)$

Fraction of transactions that contain both X and Y

$$\text{Support}(\{E, F\} \rightarrow \{B\}) = \text{support\_count}(\{B, E, F\}) / N = 2/5$$

### Confidence $P(Y|X) = P(X, Y)/P(X)$

How frequently items in Y appear in transactions that contain X

$$\begin{aligned} \text{confidence}(\{E, F\} \rightarrow \{B\}) &= \text{support}(\{B, E, F\}) / \text{support}(\{E, F\}) \\ &= \text{support\_count}(\{B, E, F\}) / \text{support\_count}(\{E, F\}) \\ &= 2/2 = 1 \end{aligned}$$

# CONFIDENCE

$$\begin{aligned}\text{confidence}(\{E,F\} \rightarrow \{B\}) \\ &= \text{support}(\{B,E,F\})/\text{support}(\{E,F\}) \\ &= \text{support\_count}(\{B,E,F\})/\text{support\_count}(\{E,F\}) \\ &= 2/2 = 1\end{aligned}$$

Switching LHS and RHS results in different rules with different confidences.