

# **Data Mining**

## **Association Rule Mining (Part A)**

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

# Association Mining: Its Definition

- Association rule mining:
  - Finding frequent patterns, associations, or correlations among sets of items or objects in transactional databases (often used for boosting sales)
- Example
  - $\text{buys}(x, \text{“diapers”}) \rightarrow \text{buys}(x, \text{“beers”}) [0.5\%, 60\%]$
  - Support = 0.5% and confidence = 60%

# An Example

- Transaction 10 buys items X,Y,Z
- Transaction 5 buys items X, Z
- Transaction 20 buys items X, W
- Transaction 40 buys items Y, M, N
- $X \rightarrow Z$  (50%, 66.6%)
- $X \rightarrow W$  (25%, 33.3%)
- $Z \rightarrow X$  (50%, 100%)
- $W \rightarrow X$  (25%, 100%)

# Important Concepts

- To calculate the support of  $X \rightarrow Z$ 
  - Calculate the support of itemset  $\{X, Z\}$ , i.e. the number of transactions buying  $X, Z$  divided by the total number of transactions
- To calculate the confidence of  $X \rightarrow Z$ 
  - Calculate the support of itemset  $\{X\}$
  - $(\text{Support of } \{X, Z\}) / (\text{Support of } \{X\})$

# Minimum Support, Confidence, Frequent Item Sets

- Transaction 10 buys items X,Y,Z
- Transaction 5 buys items X, Z
- Transaction 20 buys items X, W
- Transaction 40 buys items Y, M, N
- Minimum support = 50%
- Frequent itemsets include {X} (75%), {Y} (50%), {Z} (50%), {X, Z} (50%) but not {M} (25%), {N} (25%), {W} (25%).

# Generate All Rules

- Minimum support = 50%
- Minimum confidence = 50%
- We have frequent itemsets {A}, {B}, {C}, {E}, {B E}
- Here are all the association rules:
  - $B \rightarrow E$  (60%, 75%)
  - $E \rightarrow B$  (60%, 100%)



# End of Association Rule Mining Module (Part A)