Data Mining -- Association Rules

Instructor: Jen-Wei Huang

Office: 92528 in the EE building
jwhuang@mail.ncku

Association Rules

- Finding association, correlation or causal structures among sets of items or objects in transactional, relational DB
- Examples
  - bread ^ milk -> butter
  - age("25~35") ^ income("35,000~40,000) -> buyer(Lancer)

Definitions

- $I = \{i_1, i_2, i_3 \ldots i_n\}$: the set of all items
  - Itemset: a set of items
- Association rule: $A \rightarrow B$,
  - where $A \subset I$, $B \subset I$, $A \cap B = \varnothing$
- support $(A \rightarrow B) = Prob.(A \cup B)$
- confidence$(A \rightarrow B) = Prob.(A \cup B / A)$
  - Strong rule: satisfy both minimum support & confidence

Example

Tid	Items
100	A, C, D
200	B, C, E
300	A, B, C, E
400	B, E

min_support = 2
min_conf = 2/3

- Strong rules
  - {B, E}→C (2/3)
  - C→A (2/3)
  - A→C (2/2)

References

- Slides from Prof. J.-W. Han, UIUC
- Slides from Prof. M.-S. Chen, NTU
- Slides from Prof. W.-Z. Peng, NCTU