Lecture 4. Pattern Evaluation

Lecture 4. Pattern Evaluation

❑ Interestingness Measures in Pattern Mining

❑ Interestingness Measures: Lift and χ^2

❑ Null-Invariant Measures

❑ Comparison of Interestingness Measures

Session 1: Limitation of the Support-Confidence Framework

How to Judge if a Rule/Pattern Is Interesting?

- Pattern-mining will generate a large set of patterns/rules
  - Not all the generated patterns/rules are interesting
- Interestingness measures: Objective vs. subjective
  - Objective interestingness measures
    - Support, confidence, correlation, …
  - Subjective interestingness measures: One man’s trash could be another man’s treasure
    - Query-based: Relevant to a user’s particular request
    - Against one’s knowledge-base: unexpected, freshness, timeliness
    - Visualization tools: Multi-dimensional, interactive examination

Limitation of the Support-Confidence Framework

- Are *s* and *c* interesting in association rules: "A ⇒ B" [*s*, *c*]?

Be careful!

- Example: Suppose one school may have the following statistics on # of students who may play basketball and/or eat cereal:

	play-basketball	not play-basketball	sum (row)
eat-cereal	400	350	750
not eat-cereal	200	50	250
sum(col.)	600	400	1000

2-way contingency table

- Association rule mining may generate the following:
  - *play-basketball* ⇒ *eat-cereal* [40%, 66.7%] (higher s & c)
- But this strong association rule is misleading: The overall % of students eating cereal is 75% > 66.7%, a more telling rule:
  - ¬ *play-basketball* ⇒ *eat-cereal* [35%, 87.5%] (high s & c)