

# Finding Interesting Rules from Large Sets of Discovered Association Rules

Mika Klemettinen, Heikki Mannila, Pirjo Ronkainen, Hannu Toivonen,  
Inkeri Verkamo  
**Dept. of Computer Science**  
**University of Helsinki**

Presented By:  
Anita Lakshmanan

# PROBLEM

Hundreds of Association Rules are found by Association Rule Mining Algorithms. How do we select those that are most interesting to the user?

In effect - a “Second Order Data Mining Problem”

# Organization of Paper

- Properties of Association Rules
- Selecting Interesting Rules
- Visualization of Association Rules
- Conclusions

# Properties of AR's

- Too many rules
- Rules can correspond to prior knowledge
- Rules can refer to uninteresting (to the user) attributes
- Rules can be redundant
- Pruning based on support and confidence can cause loss of useful rules

# Approach

- Assuming rules have been found already:
- Classify attributes into a class hierarchy:

**Eg:**

{AI, C Programming, Data Communications}  $\subset$  Basic Course  $\subset$  Any Course

- Define Templates:

$A_1, \dots, A_k \Rightarrow A_{k+1}$  where A = attribute, class or expression C\* or C+;

The rule :

$B_1, \dots, B_h \Rightarrow B_{h+1}$  matches the pattern if the rule can be considered to be an instance of the pattern

**Eg:** Template: Grad Course, Any Course\*  $\Rightarrow$  Design of Algorithms

Instance : AI, Stringology  $\Rightarrow$  Design of Algorithms

# Templates

- User can specify what is interesting
- Inclusive Templates
  - Eg: Grad Course, Any Course\*  $\Rightarrow$  Design of Algorithms
  - **Using this rule - user finds all the associations that contain Design of Algorithms on the right hand side**
- Restrictive Templates
  - Eg: Basic Course, Any Course\*  $\Rightarrow$  Any course
  - **User filters out all rules that have at least one basic course on the left hand side.**

# Visualization

## Features of “Rule Visualizer” (Prototype)

### Rule Selection:

By specifying confidence  $\gamma$ , support  $\sigma$  and commonness  $\mu = \gamma * \sigma$

By entering inclusive and restrictive templates

### Rule Browsing:

View selected rules as bar graphs of support, confidence and commonness

### Rule Graphs

View several rules simultaneously - as directed hypergraphs

- Visualize rules as a hypergraph
- Attributes are the nodes
- Edge consists of all attributes appearing in a rule
- Each edge has a distinguished vertex (the attribute on the right hand side)
- Weights are the support and confidence.
- Disadvantage: graph becomes dense SOON

- Templates are used to reduce complexity
- Deleting a node = entering two restrictive templates - one with attribute on LHS and other with the attribute on the RHS
- Nodes can be marked interesting (inclusive)
- User can restrict rule sizes eg: at most 2 attributes on right hand side

# Conclusion

- Interesting rules found using class hierarchy and templates
- Domain knowledge of user, special interest areas taken into account
- Visualization tool prototype presented
- **Remarks:**
  - Redundant rules are not pruned
  - Success dependant on domain knowledge of user
  - Does not rely only on support and confidence measures for pruning
  - Efficiency of method still depends on that of the underlying algorithm used for mining AR's