

CS490D:
Introduction to Data Mining
Prof. Chris Clifton

April 19, 2004

More on Associations/Rules



Project: Association

- Several people have suggested association rule mining
 - Good idea
- Issues
 - Most data is not “itemsets”, but scaled values
 - Is apriori the right algorithm?
- Ideas
 - Try Tertius
 - Decision rules (need to define class)



Tertius

- First-order logic learning
 - Similar to ILP we have been discussing
- Provide structural schema definition
 - Individual: What is an individual (for counting purposes)
 - Structural: Define relationships between individuals
 - Properties: Things that describe an individual
- <http://www.cs.bris.ac.uk/Research/MachineLearning/Tertius/index.html>



Tertius: Use

- Call: Specify number of literals, number of variables in rules
 - Finds strongest k rules
- Predicate File
 - Parent 2 person person cwa
 - Daughter 2 person person cwa
 - Male 1 person cwa
- Fact File
 - Parent(Chris, Denise)
 - Daughter(Denise, Chris)
 - Female(Denise)



JRIP/Ripper



- Decision Rules
 - Like association rules
 - But need target class (right hand side)
- Idea:
 - Grow rule
 - Prune rule
 - If good then keep
 - Repeat
- Growth: Based on information gain
- Prune: $p+N-n / P+N$