



Classification and Prediction

——Associative Classification——

徐华

清华大学 计算机系 智能技术与系统国家重点实验室

xuhua@tsinghua.edu.cn

1

Classification and Prediction



- ◉ Basic Concepts
- ◉ Issues Regarding Classification and Prediction
- ◉ Decision Tree
- ◉ Bayesian Classification
- ◉ Neural Networks
- ◉ Support Vector Machine
- ◉ K-Nearest Neighbor
- ◉ **Associative classification**
- ◉ Classification Accuracy

2



Associative classification



◉ Associative Classification

- ◆ Association rules are generated and analyzed for classification
- ◆ Search for strong associations between frequent patterns (conjunctions of attribute-value pairs) and class labels
- ◆ Classification: Based on evaluating a set of rules in the form of

$$p_1 \wedge p_2 \dots \wedge p_l \rightarrow "A_{\text{class}} = C" \text{ (sup, conf)}$$

3



Associative classification



◉ Why effective?

- ◆ It explores highly confident associations among multiple attributes and may overcome some constraints introduced by decision-tree induction, which considers only one attribute at a time
- ◆ In many studies, associative classification has been found to be more accurate than some traditional classification methods, such as C4.5

4



Associative classification



- ◉ **CBA (Classification By Association: Liu, Hsu & Ma, KDD' 98)**
 - ◆ Mine association possible rules in the form of
 - Cond-set (a set of attribute-value pairs) → class label
 - ◆ Build classifier: Organize rules according to decreasing precedence based on confidence and then support
- ◉ **CMAR (Classification based on Multiple Association Rules: Li, Han, Pei, ICDM' 01)**
 - ◆ Classification: Statistical analysis on multiple rules
- ◉ **CPAR (Classification based on Predictive Association Rules: Yin & Han, SDM' 03)**
 - ◆ Generation of predictive rules (FOIL-like analysis)
 - ◆ High efficiency, accuracy similar to CMAR

5



Associative classification-CMAR



- ◉ **Efficiency: Uses an enhanced FP-tree that maintains the distribution of class labels among tuples satisfying each frequent itemset**
- ◉ **Rule pruning whenever a rule is inserted into the tree**
 - ◆ Given two rules, R_1 and R_2 , if the antecedent of R_1 is more general than that of R_2 and $\text{conf}(R_1) \geq \text{conf}(R_2)$, then R_2 is pruned
 - ◆ Prunes rules for which the rule antecedent and class are not positively correlated, based on a χ^2 test of statistical significance
- ◉ **Classification based on generated/pruned rules**
 - ◆ If only one rule satisfies tuple X , assign the class label of the rule
 - ◆ If a rule set S satisfies X , CMAR
 - divides S into groups according to class labels
 - uses a weighted χ^2 measure to find the strongest group of rules, based on the statistical correlation of rules within a group
 - assigns X the class label of the strongest group

6





Thanks !

