Session 3. Associative Classification: CBA and CMAR

CBA: Classification Based on Associations

- CBA [Liu, Hsu and Ma, KDD'98]
- Method
  - Mine high-confidence, high-support class association rules
  - LHS: conjunctions of attribute-value pairs); RHS: class labels
    $p_1 \wedge p_2 \ldots \wedge p_l \rightarrow$ "$A_{\text{class-label}} = C$" (confidence, support)
  - Rank rules in descending order of confidence and support
  - Classification: Apply the first rule that matches a test case; o.w. apply the default rule
  - Effectiveness: Often found more accurate than some traditional classification methods, such as C4.5
  - Why? — Exploring high confident associations among multiple attributes may overcome some constraints introduced by some classifiers that consider only one attribute at a time

CMAR: Classification Based on Multiple Association Rules

- 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 $conf(R_1) \geq conf(R_2)$, then prune R_2
  - Prunes rules for which the rule antecedent and class label are not positively correlated, based on the 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
    - Divide S into groups according to class labels
    - Use a weighted χ^2 measure to find the strongest group of rules, based on the statistical correlation of rules within a group
    - Assign X the class label of the strongest group
- CMAR improves model construction efficiency and classification accuracy