CMAR – Accurate and Efficient Classification Based on Multiple Class-Association Rules
Classification

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>4</td>
<td>5</td>
<td>B</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>5</td>
<td>A</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>7</td>
<td>C</td>
</tr>
</tbody>
</table>

Training

Classifier

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>4</th>
<th>?</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>5</td>
<td>6</td>
<td>?</td>
</tr>
</tbody>
</table>
Associative Classification

• Combine classification and pattern mining (B. Liu, W. Hsu and Y. Ma)
  - Mine Class Association Rules
    • $X,Y,Z \rightarrow C_1$
    - Build classifier
CBA – Classification Based on Association

• Mine CARs using Apriori-like approach
• Select rules for the classifier
  - Precedence order:
    • R1 has higher confidence than R2
    • R1 and R2 have same confidence but R1 has higher support
    • R1 and R2 have same confidence and support but R1 is generated first (i.e. R1 has less items than R2)
  - Single coverage
CBA – Problems

- Selection of most effective rule
  - Highest confidence
- Generation of many, many rules
  - Storage overhead (pruning)
  - Computational overhead
CMAR – Classification Based on Multiple Class Association Rules

- Set of CARs to decide the class
- FP-growth like rule mining
- CR-tree to reduce storage overhead
CMAR - Candidate generation

- Mine CARs using FP-growth adaptation
- Store CARs using CR-tree
- Pruning mechanisms
  - Precedence relationship
  - Positive correlation to class label (chi2)
  - Multiple database coverage
CMAR – Classification

- Select all rules that match object
- 2 cases:
  - All same class: OK
  - Multiple classes: use weighted chi2
Evaluation

- Outperforms C4.5 and CBA on accuracy
- Less storage requirements compared to CBA
- Lower running time compared to CBA
- Accuracy does not depend too much on confidence and coverage threshold
Still problems

- Many rules generated
- Confidence-based rule evaluation thus overfitting

⇒ CPAR: Classification based on Predictive Association Rules
Questions?