Classification by Association

Cse352
Artificial Intelligence

Professor Anita Wasilewska
Generating **Classification Rules by Association**

When mining **association rules** for use in **classification**, we are **only interested** in **association rules** of the form

\[ p_1 \land p_2 \land \ldots \land p_k \implies \text{class} = c \]

where the rule antecedent is a **conjunction** of items \( p_1, p_2, \ldots, p_k \) **associated** with a **class label** \( c \)

- The process of finding such rules is called **Classification by Association**
## Example: Original Data

<table>
<thead>
<tr>
<th>Student</th>
<th>Grade</th>
<th>Income</th>
<th>Buys</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS</td>
<td>High</td>
<td>Low</td>
<td>Milk</td>
</tr>
<tr>
<td>CS</td>
<td>High</td>
<td>High</td>
<td>Bread</td>
</tr>
<tr>
<td>Math</td>
<td>Low</td>
<td>Low</td>
<td>Bread</td>
</tr>
<tr>
<td>CS</td>
<td>Medium</td>
<td>High</td>
<td>Milk</td>
</tr>
<tr>
<td>Math</td>
<td>Low</td>
<td>Low</td>
<td>Bread</td>
</tr>
</tbody>
</table>
**STEP 1: Data Conversion**

**Converted Data**

<table>
<thead>
<tr>
<th>Student = CS (I1)</th>
<th>Student = math (I2)</th>
<th>Grade = high (I3)</th>
<th>Grade = medium (I4)</th>
<th>Grade = low (I5)</th>
<th>Income = high (I6)</th>
<th>Income = low (I7)</th>
<th>Buys = milk (I8)</th>
<th>Buys = bread (I9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>
Step 2: Apriori Algorithm
Generating 1-itemset Frequent Pattern

C1
Let, the minimum support count be 2
Since we have 5 records => minimum Support = 2/5 = 40%
Let, minimum confidence required is 70%
Generating 2-itemset Frequent Pattern

Generate C2 candidates from L1

No need of pruning here-Scan D for count of each candidate

Choose candidates with support count >= minimum support count
Generating Candidates: $C_k$

- **Join Step**: $C_k$ is generated by joining $L_{k-1}$ with itself.

- **Prune Step**: Any (k-1)-item set that is not frequent cannot be a subset of a frequent k-item set.
Example: Joining and Pruning

1. The join step: To find $C_k$, a set of candidate $k$-itemsets is generated by joining $L_{k-1}$ with itself.

$L_k$ – Itemsets       $C_k$ – Candidates

For example in our case:
Considering $\{I2,I5\}, \{I7,I9\}$ from $L_2$ to arrive at $C_3$ we Join $L_2 * L_2$

and we obtain for example $\{I2,I5,I7\}, \{I2,I5,I9\}$ as resultant candidates in $C_3$ generated from $L_2$

Considering $\{I1,I3\}, \{I1,I6\}$ from $L_2$ we generate a candidate $\{I1,I3,I6\}$ in $C_3$
Example: Joining and Pruning

2. The prune step:
$C_k$ is a superset of $L_k$, that is, its members may or may not be frequent.

$C_k$ however, can be huge and we prune it applying Apriori Principle:
“if A is a frequent item set, then each of its subsets is a frequent item set.”
It is expressed by formulation of the

Prune Step: Any (k-1)-item set that is not frequent cannot be a subset of a frequent k-item set.

Thus, $\{I2,I5,I7\}$, $\{I2,I5,I9\}$ from join step are considered since all their subsets are frequent.

but $\{I1,I3,I6\}$ is discarded since it subset $\{I3,I6\}$ is not frequent, i.e. was not in $L_2$. 
Generating 3-itemset Frequent Pattern

Generate and prune \( C_3 \) candidates from \( L_2 \)

<table>
<thead>
<tr>
<th>Item Set</th>
<th>Support Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>( {I_2,I_5,I_7} )</td>
<td>2</td>
</tr>
<tr>
<td>( {I_2,I_5,I_9} )</td>
<td>2</td>
</tr>
<tr>
<td>( {I_2,I_7,I_9} )</td>
<td>2</td>
</tr>
<tr>
<td>( {I_5,I_7,I_9} )</td>
<td>2</td>
</tr>
</tbody>
</table>

Scan \( D \) for count of candidates

Choose candidates with support count \( \geq \) min support count

<table>
<thead>
<tr>
<th>Item Set</th>
<th>Support Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>( {I_2,I_5,I_7} )</td>
<td>2</td>
</tr>
<tr>
<td>( {I_2,I_5,I_9} )</td>
<td>2</td>
</tr>
<tr>
<td>( {I_2,I_7,I_9} )</td>
<td>2</td>
</tr>
<tr>
<td>( {I_5,I_7,I_9} )</td>
<td>2</td>
</tr>
</tbody>
</table>

Generating 4-itemset Frequent Pattern

Generate and prune \( C_4 \) candidates from \( L_3 \)

<table>
<thead>
<tr>
<th>Item Set</th>
<th>Support Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>( {I_2,I_5,I_7,I_9} )</td>
<td>2</td>
</tr>
</tbody>
</table>

Scan \( D \) for count of each candidate

Choose candidates with support count \( \geq \) min support count

<table>
<thead>
<tr>
<th>Item Set</th>
<th>Support Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>( {I_2,I_5,I_7,I_9} )</td>
<td>2</td>
</tr>
</tbody>
</table>
Step3: Classification by Association

- When generating **classification by association rules**
- we are **only interested** in **association rules** of the form

\[(p_1 \land p_2 \land \ldots \land p_l) \rightarrow \text{class} = C\]

- where the rule antecedent is a **conjunction of items**
- \(p_1, p_2, \ldots, p_l\) **associated** with a **class label** \(C\)

- In our example **class is either** \(I8\) or \(I9\)
- as we want to **predict** whether a **student with given characteristics** **buys Milk** or **buys Bread**
Generating Classification Rules by Association

Let **minimum confidence** required be 70%

- **For example**, let’s consider 4-item frequent set

- $I = \{I_2, I_5, I_7, I_9\}$ where $I_9$ represents **buys-Bread**

- Its **nonempty subsets** needed to create **association rules**

- (we write \{2\} instead of \{I_2\} .. etc) are:

- \{2\}, \{5\}, \{7\}, \{9\},

- \{2,5\}, \{2,7\}, \{2,9\}, \{5,7\}, \{5,9\}, \{7,9\},

- \{2,5,7\}, \{2,5,9\}, \{2,7,9\}, \{5,7,9\}

- To create **classification rules** we consider **only** subsets that contain the **class item 9**
Generating **Classification Rules** by Association

Consider 3-itemset Frequent Sets \{2,5,9\}, \{2,7,9\}, \{5,7,9\}

We create classification by association rules as follows

- **R2**: \(5 \land 7 \rightarrow 9\) \[40\%,100\%\]
  - Confidence = \(\text{sc}\{I5,I7,I9\}/\text{sc}\{I5,I7\} = 2/2 = 100\%\)
  - **R2** is selected

- **R3**: \(2 \land 7 \rightarrow 9\) \[40\%,100\%\]
  - Confidence = \(\text{sc}\{I2,I7,I9\}/\text{sc}\{I2,I7\} = 2/2 = 100\%\)
  - **R3** is selected

- **R4**: \(2 \land 5 \rightarrow 9\) \[40\%,100\%\]
  - Confidence = \(\text{sc}\{I2,I7,I9\}/\text{sc}\{I2,I7\} = 2/2 = 100\%\)
  - **R4** is selected
Generating Classification by Association Rules

Consider 2-itemset Frequent Sets \{2,9\}, \{5,7\}, \{5,9\}, \{7,9\}, and \{1,8\} from L2

We create classification by association rules as follows

\textbf{R5} : \ 5 \rightarrow 9 \quad [40\%,100\%] \\
\quad \quad \text{Confidence} = \frac{\text{sc}\{I5,I9\}}{\text{sc}\{I9\}} = \frac{2}{2} = 100\%
\quad \quad \text{R5 is Selected}

\textbf{R6} : \ 2 \rightarrow 9 \quad [40\%,100\%] \\
\quad \quad \text{Confidence} = \frac{\text{sc}\{I2,I9\}}{\text{sc}\{I9\}} = \frac{2}{2} = 100\%
\quad \quad \text{R6 is Selected}

\textbf{R7} : \ 7 \rightarrow 9 \quad [40\%,100\%] \\
\quad \quad \text{Confidence} = \frac{\text{sc}\{I7,I9\}}{\text{sc}\{I9\}} = \frac{2}{2} = 100\%
\quad \quad \text{R7 is Selected}

\textbf{R8} : \ I1 \rightarrow I8 \quad [40\%, 66\%] \\
\quad \quad \text{Confidence} = \frac{\text{sc}\{I1,I8\}}{\text{sc}\{I1\}} = \frac{2}{3} = 66.66\%
\quad \quad \text{R8 is Rejected}
List of Selected Classification by Association Rules

- $2^5 \times 7 \rightarrow 9$ [40%,100%]
- $2^5 \rightarrow 9$ [40%,100%]
- $2^7 \rightarrow 9$ [40%,100%]
- $5^7 \rightarrow 9$ [40%,100%]
- $5 \rightarrow 9$ [40%,100%]
- $7 \rightarrow 9$ [40%,100%]
- $2 \rightarrow 9$ [40%,100%]

- We reduce the confidence to 66% to include 18
- $1 \rightarrow 8$ [40%,66%]
Test Data

<table>
<thead>
<tr>
<th>Student</th>
<th>Grade</th>
<th>Income</th>
<th>Buys</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>Low</td>
<td>Low</td>
<td>Bread</td>
</tr>
<tr>
<td>CS</td>
<td>Low</td>
<td>Low</td>
<td>Milk</td>
</tr>
<tr>
<td>Math</td>
<td>Low</td>
<td>Low</td>
<td>Milk</td>
</tr>
<tr>
<td>Math</td>
<td>Low</td>
<td>Low</td>
<td>Bread</td>
</tr>
<tr>
<td>CS</td>
<td>Medium</td>
<td>High</td>
<td>Milk</td>
</tr>
</tbody>
</table>

- **First Tuple:**
  Can be written as \( I_2 \land I_5 \land I_7 \rightarrow I_9 \)  
  [Success]
  The above rule is **correctly classified**
  And hence the Math student with low grade and low income buys bread

- **Second Tuple:**
  Can be written as \( I_1 \rightarrow I_8 \)  
  [Error]
  The above rule is **not correctly classified**

- **Third Tuple:**
  Can be written as \( I_2 \land I_5 \land I_7 \rightarrow I_8 \)  
  [Error]
  The above rule is **not correctly classified**
### Test Data

<table>
<thead>
<tr>
<th>Student</th>
<th>Grade</th>
<th>Income</th>
<th>Buys</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>Low</td>
<td>Low</td>
<td>Bread</td>
</tr>
<tr>
<td>CS</td>
<td>Low</td>
<td>Low</td>
<td>Milk</td>
</tr>
<tr>
<td>Math</td>
<td>Low</td>
<td>Low</td>
<td>Milk</td>
</tr>
<tr>
<td>Math</td>
<td>High</td>
<td>Low</td>
<td>Bread</td>
</tr>
<tr>
<td>CS</td>
<td>Medium</td>
<td>High</td>
<td>Bread</td>
</tr>
</tbody>
</table>

- **Fourth Tuple:**
  Can be written as $I_2 \land I_7 \rightarrow I_9$  
  [Success]
  The above rule is **correctly** classified
  And hence the Math student with low grade and low income buys bread

- **Fifth Tuple:**
  Can be written as $I_1 \rightarrow I_9$  
  [Success]
  The above rule is **correctly** classified

Hence we have **80% predictive accuracy**
And  **20% Error rate**