SoSe 2014: M-TANI: Big Data Analytics

Lecture 5 – 28/05/2014

Sead Izberovic

Dr. Nikolaos Korfiatis
Agenda

• Recap from the previous session
• Frequent Itemsets Mining
  ▪ Market-Basket Model
  ▪ Frequent Itemsets
  ▪ Association Rules
Introduction

• “The frequent-itemsets problem is that of finding sets of items that appear in (are related to) many of the same baskets” [1]
Market-Basket Model

- Describes a many to many mapping

- Looking for connections between items, not between baskets

from [1]
Applications

• Frequent-Itemset applications
  ▪ Plagiarism
  ▪ Biomarkers
  ▪ Related concepts

from [1]
Frequent Itemsets

• Frequent-Itemset applications
  ▪ Plagiarism
  ▪ Biomarkers
  ▪ Related concepts

from [1]
Frequent Itemsets

- Frequent-Itemset applications
  - Plagiarism
  - Biomarkers
  - Related concepts

from [1]
Frequent Itemsets

• Frequent-Itemset applications
  ▪ Plagiarism
  ▪ Biomarkers
  ▪ Related concepts

from [1]
Frequent Itemsets

• Definition
  - $I$ set of items
  - $b \in B$ is a basket
  - The *support* for $I$ is the number of baskets for which $I$ is a subset
  - $s$ is the support threshold
  - Sets are called frequent Itemsets, if $support \geq s$

from [1]
Frequent Itemsets

Example

\[ I = \{ \text{milk, coke, pepsi, beer juice} \} \]
\[ s = 3 \]

\[ b_1 = \{m, c, b\} \quad b_2 = \{m, p, j\} \]
\[ b_3 = \{m, b\} \quad b_4 = \{c, j\} \]
\[ b_5 = \{m, p, b\} \quad b_6 = \{m, c, b, j\} \]
\[ b_7 = \{c, b, j\} \quad b_8 = \{b, c\} \]

from [2]
Frequent Itemsets

Example

\[ l = \{ \text{milk, coke, pepsi, beer juice} \} \]
\[ s = 3 \]

\[ b_1 = \{m, c, b\} \quad b_2 = \{m, p, j\} \]
\[ b_3 = \{m, b\} \quad b_4 = \{c, j\} \]
\[ b_5 = \{m, p, b\} \quad b_6 = \{m, c, b, j\} \]
\[ b_7 = \{c, b, j\} \quad b_8 = \{b, c\} \]

Frequent Itemsets: \{m\}

\{m\} is a frequent itemset. \textit{support}(\{m\}) = 5 > s

from [2]
Frequent Itemsets

Example

\[ I = \{\text{milk, coke, pepsi, beer juice}\} \]
\[ s = 3 \]

\[ b_1 = \{m, c, b\} \quad b_2 = \{m, p, j\} \]
\[ b_3 = \{m, b\} \quad b_4 = \{c, j\} \]
\[ b_5 = \{m, p, b\} \quad b_6 = \{m, c, b, j\} \]
\[ b_7 = \{c, b, j\} \quad b_8 = \{b, c\} \]

Frequent Itemsets: \(\{m\}, \{c\}\)

\(\{c\}\) is a frequent itemset. \(support(\{c\}) = 5 > s\)

from [2]
Frequent Itemsets

Example

\[ l = \{milk, coke, pepsi, beer juice\} \]
\[ s = 3 \]

\[ b_1 = \{m, c, b\} \quad b_2 = \{m, p, j\} \]
\[ b_3 = \{m, b\} \quad b_4 = \{c, j\} \]
\[ b_5 = \{m, p, b\} \quad b_6 = \{m, c, b, j\} \]
\[ b_7 = \{c, b, j\} \quad b_8 = \{b, c\} \]

Frequent Itemsets: \{m\}, \{c\}, \{b\}

\{b\} is a frequent itemset. \text{support}(\{b\}) = 6 > s

from [2]
Frequent Itemsets

Example

\( I = \{milk, coke, pepsi, beer \textit{ juice}\} \)
\( s = 3 \)

\( b_1 = \{m, c, b\} \quad b_2 = \{m, p, j\} \)
\( b_3 = \{m, b\} \quad b_4 = \{c, j\} \)
\( b_5 = \{m, p, b\} \quad b_6 = \{m, c, b, j\} \)
\( b_7 = \{c, b, j\} \quad b_8 = \{b, c\} \)

Frequent Itemsets: \( \{m\}, \{c\}, \{b\}, \{j\} \)

\( \{j\} \) is a frequent itemset. \( \text{support}(\{j\}) = 4 > s \)

from [2]
Frequent Itemsets

Example

\( I = \{\text{milk, coke, pepsi, beer juice}\} \)
\( s = 3 \)

\( b_1 = \{m, c, b\} \quad b_2 = \{m, p, j\} \)
\( b_3 = \{m, b\} \quad b_4 = \{c, j\} \)
\( b_5 = \{m, p, b\} \quad b_6 = \{m, c, b, j\} \)
\( b_7 = \{c, b, j\} \quad b_8 = \{b, c\} \)

Frequent Itemsets: \( \{m\}, \{c\}, \{b\}, \{j\} \)

\( \{m, c\} \) is not a frequent itemset! \( \text{support}(\{m, c\}) = 2 < s \)

from [2]
Frequent Itemsets

Example

\[ I = \{milk, coke, pepsi, beer, juice\} \]
\[ s = 3 \]

\[ b_1 = \{m, c, b\} \]
\[ b_2 = \{m, p, j\} \]
\[ b_3 = \{m, b\} \]
\[ b_4 = \{c, j\} \]
\[ b_5 = \{m, p, b\} \]
\[ b_6 = \{m, c, b, j\} \]
\[ b_7 = \{c, b, j\} \]
\[ b_8 = \{b, c\} \]

Frequent Itemsets: \( \{m\}, \{c\}, \{b\}, \{j\} \)

\( \{m, b\} \) is a frequent itemset. \( support(\{m, b\}) = 4 > s \)

from [2]
Frequent Itemsets

Example

\( I = \{milk, coke, pepsi, beer juice\} \)
\( s = 3 \)

\( b_1 = \{m, c, b\} \quad b_2 = \{m, p, j\} \)
\( b_3 = \{m, b\} \quad b_4 = \{c, j\} \)
\( b_5 = \{m, p, b\} \quad b_6 = \{m, c, b, j\} \)
\( b_7 = \{c, b, j\} \quad b_8 = \{b, c\} \)

Frequent Itemsets: \( \{m\}, \{c\}, \{b\}, \{j\}, \{m, b\} \)

\( \{m, j\} \) is not a frequent itemset! \( \text{support}(\{m, j\}) = 2 < s \)
**Frequent Itemsets**

**Example**

\[ I = \{ \text{milk, coke, pepsi, beer juice} \} \]
\[ s = 3 \]

\[ b_1 = \{m, c, b\} \quad b_2 = \{m, p, j\} \]
\[ b_3 = \{m, b\} \quad b_4 = \{c, j\} \]
\[ b_5 = \{m, p, b\} \quad b_6 = \{m, c, b, j\} \]
\[ b_7 = \{c, b, j\} \quad b_8 = \{b, c\} \]

Frequent Itemsets: \{m\}, \{c\}, \{b\}, \{j\}, \{m, b\}

\{b, c\} is a frequent itemset! \text{support}(\{b, c\}) = 4 > s

from [2]
Frequent Itemsets

Example

\[ I = \{\text{milk, coke, pepsi, beer, juice}\} \]
\[ s = 3 \]

\[ b_1 = \{m, c, b\} \quad b_2 = \{m, p, j\} \]
\[ b_3 = \{m, b\} \quad b_4 = \{c, j\} \]
\[ b_5 = \{m, p, b\} \quad b_6 = \{m, c, b, j\} \]
\[ b_7 = \{c, b, j\} \quad b_8 = \{b, c\} \]

Frequent Itemsets: \{\text{m}\}, \{\text{c}\}, \{\text{b}\}, \{\text{j}\}, \{\text{m, b}\}, \{\text{b, c}\}

\{\text{c, j}\} \text{ is a frequent itemset! } support(\{\text{c, j}\}) = 3 = s 

from [2]
Frequent Itemsets

Example

\[l = \{\text{milk, coke, pepsi, beer juice}\}\]
\[s = 3\]

\[b_1 = \{m, c, b\}\]
\[b_2 = \{m, p, j\}\]
\[b_3 = \{m, b\}\]
\[b_4 = \{c, j\}\]
\[b_5 = \{m, p, b\}\]
\[b_6 = \{m, c, b, j\}\]
\[b_7 = \{c, b, j\}\]
\[b_8 = \{b, c\}\]

Frequent Itemsets: \{m\}, \{c\}, \{b\}, \{j\}, \{m, b\}, \{b, c\}, \{c, j\}

from [2]
Frequent Itemsets

Example

\[ I = \{ \text{milk, coke, pepsi, beer juice} \} \]
\[ s = 3 \]

\[ b_1 = \{ m, c, b \} \quad b_2 = \{ m, p, j \} \]
\[ b_3 = \{ m, b \} \quad b_4 = \{ c, j \} \]
\[ b_5 = \{ m, p, b \} \quad b_6 = \{ m, c, b, j \} \]
\[ b_7 = \{ c, b, j \} \quad b_8 = \{ b, c \} \]

Frequent Itemsets: \{ m \}, \{ c \}, \{ b \}, \{ j \}, \{ m, b \}, \{ b, c \}, \{ c, j \}

Why is \{ m, b, c \} not a frequent itemset?

from [2]
Association Rules

• “Are implications that, if a basket contains a certain set of items \{i_1, \ldots, i_n\}, then it is “likely” to contain another particular item \(j\) as well.” [1]

• The confidence of an association rule is the probability of \(j\) given \(I = \{i_1, \ldots, i_n\}\)

\[
\text{confidence}(I \rightarrow j) = \frac{\text{support}(I \cup \{j\})}{\text{support}(I)}
\]

from [1] [2]
Association Rules

• The *interest* of an association rule is the difference between its confidence and the fraction of baskets that contain \( j \)

\[
\text{interest}(I \rightarrow j) = \text{confidence}(I \rightarrow j) - \frac{|j \in B|}{|B|}
\]

• A association rule is interesting if:
  - \( \text{interest}(I \rightarrow j) > 0.5 \) or
  - \( \text{interest}(I \rightarrow j) < -0.5 \)

from [1] [2]
Association Rules

Example: Confidence and Interest

\[ b_1 = \{m, c, b\} \quad b_2 = \{m, p, j\} \]
\[ b_3 = \{m, b\} \quad b_4 = \{c, j\} \]
\[ b_5 = \{m, p, b\} \quad b_6 = \{m, c, b, j\} \]
\[ b_7 = \{c, b, j\} \quad b_8 = \{b, c\} \]

Frequent Itemsets: \{m\}, \{c\}, \{b\}, \{j\}, \{m, b\}, \{b, c\}, \{c, j\}

\[
\text{confidence}(\{m, b\} \rightarrow c) = \frac{\text{support}(\{m, b\} \cup \{c\})}{\text{support}(\{m, b\})} = \frac{2}{4} = \frac{1}{2}
\]

\[
\text{interest}(\{m, b\} \rightarrow c) = \text{confidence}(\{m, b\} \rightarrow c) - \frac{|c \in B|}{|B|} = \frac{1}{2} - \frac{5}{8} = -\frac{1}{8}
\]

from [2]
Association Rule Mining

1. Find all frequent itemsets $J$

2. Association Rule Generation
   - For every $A \subset J$, generate a rule $A \rightarrow J\setminus A$
     - Since $J$ is frequent $\rightarrow A$ is frequent
   - Compute the rule confidence
     - If a rule like $\{a, b, c\} \rightarrow d$ is below the confidence, so is $\{a, b\} \rightarrow \{c, d\}$
   - Output the rules $r \in R$, with $confidence(r) \geq ct$
     - $ct$ is a confidence threshold

from [2]
Association Rule Mining

Example with $s = 3$ and confidence $= \frac{3}{4}$

- $b_1 = \{m, c, b\}$
- $b_2 = \{m, p, j\}$
- $b_3 = \{m, b, c, n\}$
- $b_4 = \{c, j\}$
- $b_5 = \{m, p, b\}$
- $b_6 = \{m, c, b, j\}$
- $b_7 = \{c, b, j\}$
- $b_8 = \{b, c\}$

Frequent Itemsets: $\{m\}, \{b\}, \{c\}, \{j\}, \{m, b\}, \{b, c\}, \{c, j\}, \{m, c, b\}$

Association Rules:

- $\{b\} \rightarrow \{m\};$ confidence $= \frac{4}{6}$
- $\{m\} \rightarrow \{b\};$ confidence $= \frac{4}{5}$
- $\{b\} \rightarrow \{c\};$ confidence $= \frac{5}{6}$
- $\{b, m\} \rightarrow \{c\};$ confidence $= \frac{3}{4}$

from [2]
Association Rule Mining

Example with $s = 3$ and confidence $= \frac{3}{4}$

\[ b_1 = \{m, c, b\} \quad b_2 = \{m, p, j\} \]
\[ b_3 = \{m, b, c, n\} \quad b_4 = \{c, j\} \]
\[ b_5 = \{m, p, b\} \quad b_6 = \{m, c, b, j\} \]
\[ b_7 = \{c, b, j\} \quad b_8 = \{b, c\} \]

Frequent Itemsets: \{m\}, \{b\}, \{c\}, \{j\}, \{m, b\}, \{b, c\}, \{c, j\}, \{m, c, b\}

Association Rules:

\[ \{b\} \rightarrow \{m\}; \quad \text{confidence} = \frac{4}{6} \]
\[ \{m\} \rightarrow \{b\}; \quad \text{confidence} = \frac{4}{5} \]
\[ \{b\} \rightarrow \{c\}; \quad \text{confidence} = \frac{5}{6} \]
\[ \{b, m\} \rightarrow \{c\}; \quad \text{confidence} = \frac{3}{4} \]

\[ \vdots \]

from [2]
Association Rule Mining

• How to reduce the number of association rules?

1. Adjust $s$ so that we do not get too many frequent itemsets

2. Use maximum frequent itemsets
   - none of its immediate supersets is frequent

3. Use closed frequent itemsets
   - none of its immediate supersets has the same support as the itemset

from [1] [2] [3]
Association Rule Mining

Frequent Itemsets

Closed Frequent Itemsets

Maximum Frequent Itemsets

from [3]
### Association Rule Mining

Example: Maximum/Closed frequent itemsets

\[ s = 3 \]

<table>
<thead>
<tr>
<th></th>
<th>Support</th>
<th>Maximal</th>
<th>Closed</th>
</tr>
</thead>
<tbody>
<tr>
<td>{A}</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{B}</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{C}</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{A, B}</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{A, C}</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{B, C}</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{A, B, C}</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

from [2]
### Association Rule Mining

Example: Maximum/Closed frequent itemsets

\[ s = 3 \]

<table>
<thead>
<tr>
<th>Itemset</th>
<th>Support</th>
<th>Maximal</th>
<th>Closed</th>
</tr>
</thead>
<tbody>
<tr>
<td>{A}</td>
<td>4</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>{B}</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{C}</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{A, B}</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{A, C}</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{B, C}</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{A, B, C}</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Because the immediate superset \{A, B\} is frequent

---

from [2]
Example: Maximum/Closed frequent itemsets

\[ s = 3 \]

<table>
<thead>
<tr>
<th>Itemset</th>
<th>Support</th>
<th>Maximal</th>
<th>Closed</th>
</tr>
</thead>
<tbody>
<tr>
<td>{A}</td>
<td>4</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>{B}</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{C}</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{A, B}</td>
<td>4</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>{A, C}</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{B, C}</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{A, B, C}</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Because the support of the immediate superset \{A, B\} is 4

from [2]
Association Rule Mining

Example: Maximum/Closed frequent itemsets

\( s = 3 \)

<table>
<thead>
<tr>
<th>Itemset</th>
<th>Support</th>
<th>Maximal</th>
<th>Closed</th>
</tr>
</thead>
<tbody>
<tr>
<td>{A}</td>
<td>4</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>{B}</td>
<td>5</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>{C}</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{A, B}</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{A, C}</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{B, C}</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{A, B, C}</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Because the immediate superset \{B, C\} is frequent

from [2]
**Association Rule Mining**

Example: Maximum/Closed frequent itemsets

\[ s = 3 \]

<table>
<thead>
<tr>
<th>Itemset</th>
<th>Support</th>
<th>Maximal</th>
<th>Closed</th>
</tr>
</thead>
<tbody>
<tr>
<td>{A}</td>
<td>4</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>{B}</td>
<td>5</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>{C}</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{A, B}</td>
<td>4</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>{A, C}</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{B, C}</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{A, B, C}</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Because the immediate superset \{B, C\} and \{A, B\} have a smaller support.

from [2]
Association Rule Mining

Example: Maximum/Closed frequent itemsets

\[ s = 3 \]

<table>
<thead>
<tr>
<th>Itemset</th>
<th>Support</th>
<th>Maximal</th>
<th>Closed</th>
</tr>
</thead>
<tbody>
<tr>
<td>{A}</td>
<td>4</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>{B}</td>
<td>5</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>{C}</td>
<td>3</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>{A, B}</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{A, C}</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{B, C}</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{A, B, C}</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Because the immediate superset \{B, C\} is frequent

from [2]
Association Rule Mining

Example: Maximum/Closed frequent itemsets

$$s = 3$$

<table>
<thead>
<tr>
<th>Itemset</th>
<th>Support</th>
<th>Maximal</th>
<th>Closed</th>
</tr>
</thead>
<tbody>
<tr>
<td>{A}</td>
<td>4</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>{B}</td>
<td>5</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>{C}</td>
<td>3</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>{A, B}</td>
<td>4</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>{A, C}</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{B, C}</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{A, B, C}</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Because the support of immediate superset \{B, C\} is 3
**Association Rule Mining**

Example: Maximum/Closed frequent itemsets

\[ s = 3 \]

<table>
<thead>
<tr>
<th>Itemset</th>
<th>Support</th>
<th>Maximal</th>
<th>Closed</th>
</tr>
</thead>
<tbody>
<tr>
<td>{A}</td>
<td>4</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>{B}</td>
<td>5</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>{C}</td>
<td>3</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>{A, B}</td>
<td>4</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>{A, C}</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{B, C}</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{A, B, C}</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Because the immediate superset \{A, B, C\} is not frequent

from [2]
**Association Rule Mining**

Example: Maximum/Closed frequent itemsets

\[ s = 3 \]

<table>
<thead>
<tr>
<th>Itemset</th>
<th>Support</th>
<th>Maximal</th>
<th>Closed</th>
</tr>
</thead>
<tbody>
<tr>
<td>{A}</td>
<td>4</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>{B}</td>
<td>5</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>{C}</td>
<td>3</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>{A, B}</td>
<td>4</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>{A, C}</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{B, C}</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{A, B, C}</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Because the immediate superset \{A, B, C\} has a smaller support.

from [2]
### Association Rule Mining

Example: Maximum/Closed frequent itemsets

\[ s = 3 \]

<table>
<thead>
<tr>
<th></th>
<th>Support</th>
<th>Maximal</th>
<th>Closed</th>
</tr>
</thead>
<tbody>
<tr>
<td>{A}</td>
<td>4</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>{B}</td>
<td>5</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>{C}</td>
<td>3</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>{A, B}</td>
<td>4</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>{A, C}</td>
<td>2</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>{B, C}</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{A, B, C}</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Because not frequent!
### Association Rule Mining

Example: Maximum/Closed frequent itemsets

\[ s = 3 \]

<table>
<thead>
<tr>
<th>Itemset</th>
<th>Support</th>
<th>Maximal</th>
<th>Closed</th>
</tr>
</thead>
<tbody>
<tr>
<td>{A}</td>
<td>4</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>{B}</td>
<td>5</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>{C}</td>
<td>3</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>{A, B}</td>
<td>4</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>{A, C}</td>
<td>2</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>{B, C}</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>{A, B, C}</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Because the support of the immediate superset \{A, B, C\} is 2

from [2]
Association Rule Mining

Example: Maximum/Closed frequent itemsets

\[ s = 3 \]

<table>
<thead>
<tr>
<th>Itemset</th>
<th>Support</th>
<th>Maximal</th>
<th>Closed</th>
</tr>
</thead>
<tbody>
<tr>
<td>{A}</td>
<td>4</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>{B}</td>
<td>5</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>{C}</td>
<td>3</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>{A, B}</td>
<td>4</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>{A, C}</td>
<td>2</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>{B, C}</td>
<td>3</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>{A, B, C}</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Because the immediate superset \{A, B, C\} is not frequent

from [2]
## Association Rule Mining

**Example: Maximum/Closed frequent itemsets**

\( s = 3 \)

<table>
<thead>
<tr>
<th></th>
<th>Support</th>
<th>Maximal</th>
<th>Closed</th>
</tr>
</thead>
<tbody>
<tr>
<td>{A}</td>
<td>4</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>{B}</td>
<td>5</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>{C}</td>
<td>3</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>{A, B}</td>
<td>4</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>{A, C}</td>
<td>2</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>{B, C}</td>
<td>3</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>{A, B, C}</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Because the immediate superset \{A, B, C\} has a smaller support
Association Rule Mining

Example: Maximum/Closed frequent itemsets

\( s = 3 \)

<table>
<thead>
<tr>
<th></th>
<th>Support</th>
<th>Maximal</th>
<th>Closed</th>
</tr>
</thead>
<tbody>
<tr>
<td>{A}</td>
<td>4</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>{B}</td>
<td>5</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>{C}</td>
<td>3</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>{A, B}</td>
<td>4</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>{A, C}</td>
<td>2</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>{B, C}</td>
<td>3</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>{A, B, C}</td>
<td>2</td>
<td>No</td>
<td></td>
</tr>
</tbody>
</table>

Because not frequent!

from [2]
Association Rule Mining

Example: Maximum/Closed frequent itemsets

$s = 3$

<table>
<thead>
<tr>
<th>Itemset</th>
<th>Support</th>
<th>Maximal</th>
<th>Closed</th>
</tr>
</thead>
<tbody>
<tr>
<td>${A}$</td>
<td>4</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>${B}$</td>
<td>5</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>${C}$</td>
<td>3</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>${A, B}$</td>
<td>4</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>${A, C}$</td>
<td>2</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>${B, C}$</td>
<td>3</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>${A, B, C}$</td>
<td>2</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Because no immediate superset has the same support

from [2]
Literature

   Mining of Massive Datasets  
   Cambridge University Press

   Slides: Mining Massive Data Sets  
   URL: http://www.stanford.edu/class/cs246/slides/02-assocrules.pdf

   Slides: Introduction to Data Mining  