Efficient Hiding of Collaborative Recommendation Association Rules with Updates

Shyue-Liang Wang¹, Ting-Zheng Lai³, Tzung-Pei Hong², Yu-Lung Wu³

National University of Kaohsiung¹,²
I-Shou University³,⁴
Kaohsiung, Taiwan
Privacy Preserving Data Mining

- Privacy Preserving
  - Given $D_O$
  - Minimum support
  - Minimum confidence
  - $R_H$
    - Rules to be hidden
  - Find $D_M$ s.t. $R_M \cong R_O - R_H$
    - $D_M$: modified database
    - $R_M$: modified rules
Problem Description

- Collaborative recommendation association rules
  - \( R_3, R_5 \) same prediction as \( R_1 \) to \( R_9 \)
- Input: \( D_O, \) min_supp, min_conf, \( C \)
- Output: \( R_O \)

<table>
<thead>
<tr>
<th>TID</th>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>ABC</td>
</tr>
<tr>
<td>T2</td>
<td>ABC</td>
</tr>
<tr>
<td>T3</td>
<td>ABC</td>
</tr>
<tr>
<td>T4</td>
<td>AB</td>
</tr>
<tr>
<td>T5</td>
<td>A</td>
</tr>
<tr>
<td>T6</td>
<td>AC</td>
</tr>
</tbody>
</table>

\( D_O: \{R_3, R_5\} \)

\( \text{min_supp}=33\% \)
\( \text{min_conf}=70\% \)

\( |A|=6, |B|=4, |C|=4 \)
\( |AB|=4, |AC|=4, |BC|=3 \)
\( |ABC|=3 \)

\begin{table}
\begin{tabular}{|c|c|c|}
\hline
\textbf{I} & \textbf{J} & \textbf{Conf} \\
\hline
1 & B \Rightarrow A & (66\%, 100\%) \\
2 & C \Rightarrow A & (66\%, 100\%) \\
3 & B \Rightarrow C & (50\%, 75\%) \\
4 & C \Rightarrow B & (50\%, 75\%) \\
5 & AB \Rightarrow C & (50\%, 75\%) \\
6 & AC \Rightarrow B & (50\%, 75\%) \\
7 & BC \Rightarrow A & (50\%, 100\%) \\
8 & C \Rightarrow AB & (50\%, 75\%) \\
9 & B \Rightarrow AC & (50\%, 75\%) \\
10 & A \Rightarrow B & (66\%, 66\%) \\
11 & A \Rightarrow C & (66\%, 66\%) \\
12 & A \Rightarrow BC & (50\%, 50\%) \\
\hline
\end{tabular}
\end{table}

Not AR

\( R_O: \{R_3, R_5\} \)
Problem Description

• Input: $D_O$, $H$ (items to be hidden on RHS), min_supp, min_conf

• Output: $D_M$, $R_M$

$H = \{C\}$
min_supp = 33%
min_conf = 70%

$D_O, D_M$

<table>
<thead>
<tr>
<th>TID</th>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>ABC</td>
</tr>
<tr>
<td>T2</td>
<td>ABC</td>
</tr>
<tr>
<td>T3</td>
<td>ABC</td>
</tr>
<tr>
<td>T4</td>
<td>AB</td>
</tr>
<tr>
<td>T5</td>
<td>A</td>
</tr>
<tr>
<td>T6</td>
<td>AC</td>
</tr>
</tbody>
</table>

$|A|=6, |B|=4, |C|=3$
$|AB|=4, |AC|=3, |BC|=2, |ABC|=2$

$R_M : \{R1, R2, R7\}$
Problem Description

(1) Combine, hide (DCBS)

(2) Hide, combine (MSCR)
Numerical Experiments

Multiple Updates

Data Size

Time Effects
Numerical Experiments

Database Effects

Percentage

0% 2% 4% 6% 8% 10%

Data Size

10k 15K 20K 25K

Database Effects

DCBS
MSCR
Numerical Experiments

Multiple Updates Side Effects

Percentage

New Rules

Lost Rules

Hiding Failure

Data Size

10K 15K 20K 25K

Side Effects for MSCR
Numerical Experiments

DCBS Side Effects

Data Size

Percentage

New Rules
Lost Rules
Hiding Failure

10K 15K 20K 25K

Side Effects for DCBS
Conclusion

- Study the problem of hiding collaborative recommendation association rules with updates
- Automatically hide collaborative recommendation association rule sets without pre-mining and selection of a class of rules
- MSCR faster than DCBS, with similar side effects