Multivalued Dependencies

In relation $R$, we say MVD $X \rightarrow Y$ holds if whenever there are tuples $s$ and $t$ in $R$ such that $\pi_X(s) = \pi_X(t)$, then there is a tuple $r$ in $r$ such that:

1. $\pi_{XY}(r) = \pi_{XY}(s)$.
2. $\pi_{[R \setminus Y] \cup X}(r) = \pi_{[R \setminus Y] \cup X}(t)$.

- I.e., $r$ agrees with $s$ on the attributes mentioned, and with $t$ on $X$ and all the attributes not mentioned.

Example

Consider $CHRSG$. In addition to the FD’s $C \rightarrow T$; $HR \rightarrow C$; $HT \rightarrow R$; $HS \rightarrow R$; $CS \rightarrow G$; $CH \rightarrow R$, we might reasonably expect the MVD $C \rightarrow HR$.

- That is, given a course, the hour-room pairs are independent of the teacher-student-grade triples.
  - There will be a unique teacher for the course, but student-grade information should appear in combination with each hour-room pair, since there is no logical reason to assign different grades for different hours and rooms.

Axiomatization of FD’s + MVD’s

Start with Armstrong’s Axioms. Then add:

A4: Complementation. If $X \rightarrow Y$ holds in $R$, then $X \rightarrow (R \setminus X \setminus Y)$.

A5: MVD augmentation. If $X \rightarrow Y$, then $XZ \rightarrow YZ$ for any set of attributes $Z$.

A6: MVD transitivity. If $X \rightarrow Y$ and $Y \rightarrow Z$, then $X \rightarrow (Z \setminus Y)$.

A7: Promotion. If $X \rightarrow Y$, then $X \rightarrow Y$.

A8: If $X \rightarrow Y$ and $U \rightarrow V$, where:
  1. $U$ is disjoint from $Y$, and
  2. $V \subseteq Y$,
  then $X \rightarrow V$.

- Example: $AB \rightarrow CDE$ and $FG \rightarrow C$ imply $AB \rightarrow C$. 
Decomposition Rules

- For FD’s, we know that if $X \rightarrow A_1 \cdots A_n$, then we can “decompose” into $X \rightarrow A_1, \ldots, X \rightarrow A_n$.
- MVD’s do not always allow right sides of size 1.
- However, $X \rightarrow (R - X)$ always holds by A1 (trivial FD’s), A7, and A4. We can break $R - X$ into some disjoint partition, say

$$X \rightarrow Y_1 | Y_2 | \cdots | Y_k$$

such that $X \rightarrow Z$ iff $Z$ is a union of some of the $Y_i$’s.

Example

For $CTHRSG$, we have

$$C \rightarrow T | HR | SG$$

i.e., a course has a teacher (only one because of the FD $C \rightarrow T$), a set of hour-room pairs, and a set of student-grade pairs.

Generalized Dependencies

Unify FD’s, MVD’s, lossless joins, and inferences about dependencies. A generalized dependency consists of:

1. One or more hypothesis rows consisting of abstract symbols, one for each attribute of the relation in question.
2. A conclusion that is either another row or an equality between two symbols.
   - If the conclusion is a row, then the GD is a tuple-generating dependency (TGD); if an equality it is an equality-generating dependency (EGD).

- We shall usually talk about typed dependencies, which means that a symbol may not appear in two different columns, and we may not equate (in the conclusion) symbols from different columns.
- Usually, we talk about full dependencies, meaning that the conclusion row of a TGD uses only symbols that have appeared elsewhere.
• But sometimes we like to have partial
dependencies, where the conclusion introduces
new symbols.

Semantics of GD’s

If there is a mapping from the symbols of the
hypothesis rows that turns each hypothesis into
a tuple of R, then the conclusion must hold.

• If it is an EGD, then R must be such that
the equated symbols are mapped to the same
value.

• If a full TGD, then the conclusion row, with
the same mapping applied, must also be a row
of R.

• If a partial TGD, then there is a way to map
the new symbols in the conclusion so the
conclusion row becomes a tuple of R.

Examples

• In $R = ABCD$ the FD $AB \rightarrow C$:

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>$b_1$</td>
<td>$c_1$</td>
<td>$d_1$</td>
</tr>
<tr>
<td>$a_1$</td>
<td>$b_1$</td>
<td>$c_2$</td>
<td>$d_2$</td>
</tr>
</tbody>
</table>

$c_1 = c_2$

• The MVD $A \rightarrow BC$:

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>$b_1$</td>
<td>$c_1$</td>
<td>$d_1$</td>
</tr>
<tr>
<td>$a_1$</td>
<td>$b_2$</td>
<td>$c_2$</td>
<td>$d_2$</td>
</tr>
<tr>
<td>$a_1$</td>
<td>$b_1$</td>
<td>$c_1$</td>
<td>$d_2$</td>
</tr>
</tbody>
</table>

• The join dependency (JD) $\triangleright (AB, BC, CD)$,
which says that $AB$, $BC$, and $CD$ are a
lossless-join decomposition of $ABCD$:

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>$b$</td>
<td>$c_1$</td>
<td>$d_1$</td>
</tr>
<tr>
<td>$a_2$</td>
<td>$b$</td>
<td>$c$</td>
<td>$d_2$</td>
</tr>
<tr>
<td>$a_3$</td>
<td>$b_3$</td>
<td>$c$</td>
<td>$d$</td>
</tr>
</tbody>
</table>

The Chase Infers Full GD’s

Given a set of full GD’s $\mathcal{G}$, does another GD $G$
follow? (I.e., does every relation instance that
satisfies $\mathcal{G}$ also satisfy $G$?)
Start with a relation $R$ consisting of the hypothesis rows of $G$.

Repeatedly apply GD's $H$ of $G$ to the current relation:

a) If $H$ is a TGD, map its hypothesis rows to tuples of $R$ in any way, and insert the mapped conclusion row of $H$ into $R$.

b) If $H$ is an EGD, map its hypothesis rows to tuples of $R$ in any way and equate the symbols of $R$ that $H$'s conclusion says are equal.

- Remember to equate all occurrences of the symbols that the EGD says are equal.
- Subtle point: if one or both of the equated symbols appear in the conclusion, change these occurrences as well; i.e., the desired conclusion must change as well as the tuples in the constructed relation.

Since no new symbols are ever generated (because of the “full” assumption), eventually this process stalls.

At that time, if the conclusion row of $G$ has been added to $R$ (in the case $G$ is a TGD) or the symbols that $G$ says must be equal have been equated (in case $G$ is an EGD), then conclude that $G$ follows from $G$. If not, then not.

Proof the Chase Works

- Soundness: Each inference made is sound, since it is a direct application of a given GD.
- Completeness: Suppose the conclusion of $G$ is not obtained. Then the final $R$ is a counterexample:
  - It satisfies every GD in $G$, but does not satisfy $G$.

Proof That a Key Plus Relations From a Minimal Cover Have a Lossless Join

Suppose we have a minimal cover $\mathcal{F}$ consisting of FD's $X_i \rightarrow A_i$ for $i = 1, 2, \ldots, n$, and we choose database schema

$$\{X_1A_1, \ldots, X_nA_n, K\}$$

where $K$ is a key for the entire relation $R$. 

• We need to infer from $F$ (written as EGD’s, if you like) the JD that says the join is lossless. This JD has:

1. A conclusion row that has “unsubscripted” symbols corresponding to each attribute.

2. For each relation schema, a hypothesis row that has the unsubscripted symbol for an attribute if that attribute is in the schema, and a unique symbol $b_j$ if $b$ is the symbol for the attribute and $j$ is the row number.

---

Example

Let $F = \{A \rightarrow B, C \rightarrow D\}$. Recall the generated decomposition is $\{AB, AC, CD\}$. The JD is:

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$a$</td>
<td>$b$</td>
<td>$c_1$</td>
<td>$d_1$</td>
</tr>
<tr>
<td>2</td>
<td>$a$</td>
<td>$b_2$</td>
<td>$c$</td>
<td>$d_2$</td>
</tr>
<tr>
<td>3</td>
<td>$a_3$</td>
<td>$b_3$</td>
<td>$c$</td>
<td>$d$</td>
</tr>
</tbody>
</table>

---

Proof That This JD Follows From the FD’s

• We know $K^+ = R$.

• An induction on the order in which attributes are added to $K^+$ says that in the row for $K$, each subscripted symbol is equated to its unsubscripted version.

  • Key idea: if we use $X \rightarrow A$ to add $A$ to $K^+$, then in the row for $K$, all symbols for the attributes in $X$ have lost their subscripts. Therefore, we may apply $X \rightarrow A$ to the rows for $K$ and $XA$ to infer that $a_j = a$, where $j$ is the row number for $XA$.

• As a result, the row for $K$ eventually loses all its subscripts, and becomes the conclusion row of the JD.