Multivalued Dependencies

In relation R, we say MVD $X \twoheadrightarrow Y$ 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. $\pi_{(R-Y)\cup X}(r) = \pi_{(R-Y)\cup X}(t)$.

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

Example

Consider $CTHRSG$. 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 $C \twoheadrightarrow HR$.

- 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 $X \twoheadrightarrow Y$ holds in R, then $X \twoheadrightarrow (R - X - Y)$.

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

A6: *MVD transitivity*. If $X \twoheadrightarrow Y$ and $Y \twoheadrightarrow Z$, then $X \twoheadrightarrow (Z - Y)$.

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

A8: If $X \twoheadrightarrow Y$ and $U \rightarrow V$, where:

1. U is disjoint from Y, and

2. $V \subseteq Y$,

then $X \rightarrow V$.

❖ Example: $AB \twoheadrightarrow CDE$ 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, $X \rightarrow\rightarrow (R - X)$ always holds by A1 (trivial FD's), A7, and A4. We can break $R - X$ into *some* disjoint partition, say

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

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

Example

For $CTHRSG$, we have

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

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 $AB \to C$:

A	B	C	D
a_1	b_1	c_1	d_1
a_1	b_1	c_2	d_2
			$c_1 = c_2$

- The MVD $A \twoheadrightarrow BC$:

A	B	C	D
a_1	b_1	c_1	d_1
a_1	b_2	c_2	d_2
a_1	b_1	c_1	d_2

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

A	B	C	D
a	b	c_1	d_1
a_2	b	c	d_2
a_3	b_3	c	d
a	b	c	d

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 $\mathcal{G}$ 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 $\mathcal{G}$, 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 $X_i \to A_i$ 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 $\mathcal{F}$ (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 $\mathcal{F} = \{A \to B, C \to D\}$. Recall the generated decomposition is $\{AB, AC, CD\}$. The JD is:

A	B	C	D
a	b	c_1	d_1
a	b_2	c	d_2
a_3	b_3	c	d
a	b	c	d

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 $X \to A$ 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 $X \to A$ 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.