On the logical Implication of Multivalued Dependencies with Null Values

Sebastian Link

Information Science Research Centre
Dept of Information Systems, Massey University
New Zealand

This research is supported by Marsden Funding, Royal Society of New Zealand.

Relational Databases with Null Values

- schema: finite set R of attributes (with domains), e.g.

$$\textsc{Work}=\{\text{Employee, Child, Salary, Year}\}$$

- database: partial relations r, e.g.

Employee	**Child**	**Salary**	**Year**
Homer	**Bart**	**2000**	**2005**
Homer	**Lisa**	**2200**	**2006**
Homer	**Bart**	**2200**	**2006**
Homer	**Lisa**	**2000**	**2005**
Mr Burns	ν	**8000**	**2005**
Mr Burns	ν	**9000**	**2006**

- attributes: A with domains $dom(A)$
- finite sets of attributes: U, V, X, Y, Z, X_i, Y_i $(\subseteq R)$
- $r[Y], r_X[Y], Dom(r)$

Multivalued Dependencies

- NMVDs: expressions $X \twoheadrightarrow Y$, e.g., Employee $\twoheadrightarrow$ Child
- $\models_r X \twoheadrightarrow Y$ precisely when the X-total subrelation of r is decomposable into two of its projections without loss of information

$$r_X[R] = r_X[XY] \bowtie r_X[X(R-Y)]$$

- MVD Employee $\twoheadrightarrow$ Child is satisfied by previous database
- store independent facts separately (less redundancy, better updating)

Employee	Child
Homer	Bart
Homer	Lisa
Mr Burns	ν

Employee	Salary	Year
Homer	2000	2005
Homer	2200	2006
Mr Burns	8000	2005
Mr Burns	9000	2006

The Standard Notion of R-implication

- $\Sigma = \{X_1 \twoheadrightarrow Y_1, \ldots, X_k \twoheadrightarrow Y_k\}$, $X \twoheadrightarrow Y$ on R $(X \cup Y \cup \bigcup_{i=1}^{k}(X_i \cup Y_i) \subseteq R)$
  Σ R-implies $X \twoheadrightarrow Y$ iff $\forall r \subseteq dom(R)$: if $\models_r \Sigma$, then $\models_r X \twoheadrightarrow Y$

$$\frac{}{X \twoheadrightarrow Y}\, Y \subseteq X \qquad \frac{X \twoheadrightarrow Y}{XU \twoheadrightarrow YV}\, V \subseteq U \qquad \frac{X \twoheadrightarrow Y}{X \twoheadrightarrow R - Y}$$

(reflexivity, $\mathcal{R}$) (augmentation, $\mathcal{A}$) (R-complementation, $\mathcal{C}_R$)

$$\frac{X \twoheadrightarrow Y, X \twoheadrightarrow Z}{X \twoheadrightarrow YZ} \qquad \frac{X \twoheadrightarrow Y, X \twoheadrightarrow Z}{X \twoheadrightarrow Z - Y} \qquad \frac{X \twoheadrightarrow Y, X \twoheadrightarrow Z}{X \twoheadrightarrow Y \cap Z}$$

(union, $\mathcal{U}$) (difference, $\mathcal{D}$) (intersection, $\mathcal{I}$)

- Lien'82: $R\mathfrak{K} = \langle \mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{C}_R \rangle$ R-complete ($\forall \Sigma$ on R: $\Sigma^+_{R\mathfrak{K}} = \Sigma^*_R$)
- $R\mathfrak{K}_1 = \langle \mathcal{R}, \mathcal{A}, \mathcal{I}, \mathcal{C}_R \rangle$ and $R\mathfrak{K}_2 = \langle \mathcal{R}, \mathcal{A}, \mathcal{D}, \mathcal{C}_R \rangle$ also R-complete

The Role of the Complementation Rule

- complementation rule $\mathcal{C}_R$ enjoys special status
- Does Employee $\twoheadrightarrow$ Child imply Employee $\twoheadrightarrow$ Salary?
- the answer depends on the underlying relation schema R
- on $R = \{\text{Employee, Child, Salary}\}$: yes!
- on $R = \{\text{Employee, Child, Salary, Year}\}$: no!
- $\mathcal{C}_R$ just a means of database normalisation on universe R
- two ways to go:

(i) the status of $\mathcal{C}_R$ should be reflected within axiomatisation
(ii) find notion of implication independent from underlying schema

Complementary Axiomatisations

- $R\mathfrak{S} = \mathfrak{S} \cup \{\mathcal{C}_R\}$ said to be R-*complementary*:

  every inference of $X \twoheadrightarrow Y$ using $R\mathfrak{S}$ can be turned into inference of $X \twoheadrightarrow Y$ using $R\mathfrak{S}$ such that R-complementation rule $\mathcal{C}_R$ applied at most once and if, then in the last step

- formally:

$$X \twoheadrightarrow Y \in \Sigma^+_{R\mathfrak{S}} \quad \text{iff} \quad X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{S}} \text{ or } X \twoheadrightarrow (R - Y) \in \Sigma^+_{\mathfrak{S}}$$

  where $\Sigma = \{X_1 \twoheadrightarrow Y_1, \ldots, X_k \twoheadrightarrow Y_k\}$ and $X \cup Y \cup \bigcup_{i=1}^{k} (X_i \cup Y_i) \subseteq R$

- $R\mathfrak{S}$ said to be *sound* (*complete*, *complementary*) iff $R\mathfrak{S}$ is R-sound (R-complete, R-complementary) for all R
- $R\mathfrak{K}$, $R\mathfrak{K}_1$, $R\mathfrak{K}_2$ are all not complementary

Why $R\mathfrak{K}$ isn't complementary

- $\Sigma = \{\text{Employee} \twoheadrightarrow \text{Child Salary}, \text{Employee} \twoheadrightarrow \text{Child}\}$
- $\text{Employee} \twoheadrightarrow \text{Salary} \notin \Sigma^{+}_{\{\mathcal{R},\mathcal{A},\mathcal{U}\}}$
- $\text{Employee} \twoheadrightarrow Y \notin \Sigma^{+}_{\{\mathcal{R},\mathcal{A},\mathcal{U}\}}$

$$\forall Y.Y - \{\text{Employee, Child, Salary}\} \neq \emptyset$$

- for $R := \{\text{Employee, Child, Salary, Year}\}$ we have

$$\text{Employee} \twoheadrightarrow \text{Salary} \in \Sigma^{+}_{R\mathfrak{K}}$$

- in any such inference $\mathcal{C}_R$ must be used at least once, but $R - \{\text{Salary}\} = \{\text{Employee, Child, Year}\}$ implies that $\mathcal{C}_R$ is not just used as last rule

Example Derivation

- inference of Employee $\twoheadrightarrow$ Salary from Employee $\twoheadrightarrow$ Child, Salary and Employee $\twoheadrightarrow$ Child using $\langle \mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{C}_R \rangle$

$$\cfrac{\cfrac{\cfrac{\text{Employee} \twoheadrightarrow \text{Child, Salary}}{\text{Employee} \twoheadrightarrow \text{Employee, Year}}\,\mathcal{C}_R \qquad \text{Employee} \twoheadrightarrow \text{Child}}{\text{Employee} \twoheadrightarrow \text{Employee, Child, Year}}\,\mathcal{U}}{\text{Employee} \twoheadrightarrow \text{Salary}}\,\mathcal{C}_R$$

- the difference rule $\mathcal{D}$ allows us to infer the NMVD in one single step

Complementary Axiomatisation for NMVDs

- are there complementary axiomatisations at all?
- Theorem:
  $R\mathfrak{L} = \langle \mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{D}, \mathcal{C}_R \rangle$ is complete and complementary for the R-implication of NMVDs
- Theorem:
  The only minimal subset of $\langle \mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{D}, \mathcal{I}, \mathcal{C}_R \rangle$ which is both complete and complementary is indeed $R\mathfrak{L}$.
- The system $\langle \mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{D} \rangle$ is nearly complete for R-implication.

Implication in undetermined Universes

- consequences dependent on the universe are in fact no consequences
- expression: $X \twoheadrightarrow Y$ with finite X, Y
- partial r *satisfies* $X \twoheadrightarrow Y$ ($\models_r X \twoheadrightarrow Y$) iff

  $X \cup Y \subseteq Dom(r)$ and $r_X[Dom(r)] = r_X[XY] \bowtie r_X[X \cup (Dom(r) - Y)]$
- $\Sigma = \{X_1 \twoheadrightarrow Y_1, \ldots, X_k \twoheadrightarrow Y_k\} \models X \twoheadrightarrow Y$ iff

  for each partial relation r with $X \cup Y \cup \bigcup_{i=1}^{k}(X_i \cup Y_i) \subseteq Dom(r)$ we have: if $\models_r \Sigma$, then $\models_r X \twoheadrightarrow Y$
- $X \cup Y \cup \bigcup_{i=1}^{k}(X_i \cup Y_i) \subseteq R$:

  Σ R-implies $X \twoheadrightarrow Y$ whenever Σ implies $X \twoheadrightarrow Y$, but not vice versa!

Axiomatising NMVDs in undetermined Universes

- Theorem:
  The system $\mathfrak{L} = \langle \mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{D} \rangle$ is sound and complete for the implication of NMVDs. (for all finite Σ we have $\Sigma^+_{\mathfrak{L}} = \Sigma^*$)

- Theorem:
  $\mathfrak{L}$ is the only minimal subset of $\langle \mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{D}, \mathcal{I} \rangle$ that is complete for the implication of NMVDs.

- Theorem (Minimising Minimality):
  The following system is complete for the implication of NMVDs

$$\frac{}{\emptyset \twoheadrightarrow \emptyset} \quad \frac{}{A \twoheadrightarrow A} \quad \frac{X \twoheadrightarrow Y}{XA \twoheadrightarrow Y} \quad \frac{X \twoheadrightarrow Y, X \twoheadrightarrow Z}{X \twoheadrightarrow YZ} \quad \frac{X \twoheadrightarrow Y, X \twoheadrightarrow Z}{X \twoheadrightarrow Z - Y}$$

Correspondences between Axiomatisations

- Theorem:
  Let $\mathfrak{S}$ be a sound set of inference rules for the implication of NMVDs. The set $\mathfrak{S}$ is complete for the implication of NMVDs if and only if $R\mathfrak{S}$ is complete and complementary for the R-implication of NMVDs.

- Theorem:
  Let $\mathfrak{S}$ be a sound set of inference rules for the implication of NMVDs. The set $\mathfrak{S}$ is minimal and complete for the implication of NMVDs if and only if $R\mathfrak{S}$ is complete and complementary for the R-implication of NMVDs, and there is no inference rule $\Re \in R\mathfrak{S}$ such that the set $R(\mathfrak{S} - \{\Re\})$ is still both complete and complementary for the R-implication of NMVDs.

A strong Notion of Minimality

- $\mathfrak{R}$ is *R-strongly independent* from $R\mathfrak{S}$ iff there is some Σ on R and some **trivial** σ on R such that $\sigma \notin \Sigma^+_{R\mathfrak{S}}$ but $\sigma \in \Sigma^+_{R\mathfrak{S}\cup\{\mathfrak{R}\}}$
- $\mathfrak{R}$ *strongly independent* from $R\mathfrak{S}$ iff there is some R such that $\mathfrak{R}$ is R-strongly independent from $R\mathfrak{S}$
- $R\mathfrak{S}$ *strongly minimal* iff for all $\mathfrak{R} \in R\mathfrak{S}$: $\mathfrak{R}$ strongly independent from $R\mathfrak{S} - \{\mathfrak{R}\}$
- $R\mathfrak{K}$, $R\mathfrak{K}_1$, $R\mathfrak{K}_2$ are not strongly minimal
- Theorem:
  The following system is strongly minimal for NMVD R-implication.

$$\frac{}{A \twoheadrightarrow A} \qquad \frac{}{\emptyset \twoheadrightarrow R} \qquad \frac{X \twoheadrightarrow Y}{XA \twoheadrightarrow Y} \qquad \frac{X \twoheadrightarrow Y, X \twoheadrightarrow Z}{X \twoheadrightarrow Z - Y}$$

Interesting Things to look at

(i) Are there any minimal sets of inference rules that are also complementary?

(ii) What is the time-complexity of related implication problems?

(iii) Further investigate the notion of *strong minimality*, e.g. strongly minimal axiomatisations for FDs, MVDs, NFDs, NMVDs.

(iv) Consider complete axiomatisations of MVDs in Entity-Relationship Models. Are there complete axiomatisations over undetermined universes?

(v) Consider complete axiomatisations of MVDs in Nested Database Models. Are these complementary?

(vi) Consider complete axiomatisations of fuzzy and approximate MVDs. Are these complementary?