A strongly minimal axiomatisation of multivalued dependencies in incomplete database relations

This talk is dedicated to my father, Hans-Jürgen Link, who turns 62 today

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.

1. MVDs in total and partial relations
2. Classical Results
3. A new axiomatisation
4. Strong minimality
5. Conclusion and Future Work

MVDs in total relations

- DVD={Title,Actor,Role,Feature} with Title $\twoheadrightarrow$ Actor, Role

Title	Actor	Role	Feature
The Godfather	Marlon Brando	Don Vito Corleone	Deleted Scene
The Godfather	Al Pacino	Michael Corleone	Making of
The Godfather	Marlon Brando	Don Vito Corleone	Making of
The Godfather	Al Pacino	Michael Corleone	Deleted Scence
Goldmember	Mike Myers	Austin Powers	Picture Gallery
Goldmember	Mike Myers	Dr. Evil	Subtitle
Goldmember	Mike Myers	Fat Bastard	Music Video
Goldmember	Mike Myers	Austin Powers	Subtitle
Goldmember	Mike Myers	Dr. Evil	Music Video
Goldmember	Mike Myers	Fat Bastard	Picture Gallery
Goldmember	Mike Myers	Austin Powers	Music Video
Goldmember	Mike Myers	Dr. Evil	Picture Gallery
Goldmember	Mike Myers	Fat Bastard	Subtitle

- r satifies $X \twoheadrightarrow Y$ if and only if $r = r[XY] \bowtie r[X(R - Y)]$
- Σ R-implies φ iff $\forall r \subseteq dom(R)$ if r satisfies all $\sigma \in \Sigma$, then also φ

Decomposition of Example

- Title $\twoheadrightarrow$ Actor, Role separates {Title,Actor,Role,Feature} into {Title,Actor,Role} and {Title,Feature}

Title	Actor	Role
The Godfather	Marlon Brando	Don Vito Corleone
The Godfather	Al Pacino	Michael Corleone
Goldmember	Mike Myers	Austin Powers
Goldmember	Mike Myers	Dr. Evil
Goldmember	Mike Myers	Fat Bastard

Title	Feature
The Godfather	Deleted Scene
The Godfather	Making of
Goldmember	Picture Gallery
Goldmember	Subtitle
Goldmember	Music Video

NMVDs in partial relations

- t_1 *subsumes* t_2 if $\forall A \in R$: $t_1[A] = t_2[A]$ or $t_2[A] = \nu$
- no relation contains two tuples t_1 and t_2 such that t_1 subsumes t_2
- WORK={Employee, Child, Salary, Year} with Employee $\twoheadrightarrow$ Child

Employee	**Child**	**Salary**	**Year**
ν	**Maggie**	**3000**	**2005**
ν	**Lisa**	**3300**	**2006**
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**

- r satisfies $X \twoheadrightarrow Y$ iff $r_X[R] = r_X[XY] \bowtie r_X[X(R-Y)]$ where $r_X[Z] = \{t \in r[Z] \mid t \text{ is } X\text{-total}\}$

Decomposition of Example

- Employee $\twoheadrightarrow$ Salary, Year separates {Employee, Child, Salary, Year} into {Employee, Child} and {Employee,Salary,Year}

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

Inference Rules for MVDs

$$\frac{}{X \twoheadrightarrow Y}\; Y \subseteq X$$
(reflexivity, $\mathcal{R}$)

$$\frac{X \twoheadrightarrow Y}{XU \twoheadrightarrow YV}\; V \subseteq U$$
(augmentation, $\mathcal{A}$)

$$\frac{X \twoheadrightarrow Y, Y \twoheadrightarrow Z}{X \twoheadrightarrow Z - Y}$$
(pseudo-transitivity, $\mathcal{T}$)

$$\frac{}{X \twoheadrightarrow A}\; A \in X$$
(membership, $\mathcal{M}$)

$$\frac{X \twoheadrightarrow Y}{X \twoheadrightarrow R - Y}$$
(R-complementation, $\mathcal{C}_R$)

$$\frac{}{\emptyset \twoheadrightarrow R}$$
(R-axiom, $\mathcal{C}.1$)

$$\frac{X \twoheadrightarrow Y, X \twoheadrightarrow Z}{X \twoheadrightarrow YZ}$$
(union, $\mathcal{U}$)

$$\frac{X \twoheadrightarrow Y, X \twoheadrightarrow Z}{X \twoheadrightarrow Z - Y}$$
(difference, $\mathcal{D}$)

$$\frac{X \twoheadrightarrow Y, X \twoheadrightarrow Z}{X \twoheadrightarrow Y \cap Z}$$
(intersection, $\mathcal{I}$)

- Beeri, Fagin, Howard: $R\mathfrak{F} = \langle \mathcal{R}, \mathcal{A}, \mathcal{T}, \mathcal{C}_R \rangle$
- Mendelzon: $R\mathfrak{M} = \langle \mathcal{R}, \mathcal{T}, \mathcal{C}_R \rangle$ and Biskup: $R\mathfrak{B} = \langle \mathcal{C}.1, \mathcal{A}, \mathcal{T} \rangle$
- Hartmann/Link: $\langle \mathcal{C}.1, \mathcal{M}, \mathcal{T} \rangle$ and exactly one of $\{\mathcal{U}, \mathcal{D}, \mathcal{I}\}$

Inference Rules for NMVDs

$$\frac{}{X \twoheadrightarrow Y}\ Y \subseteq X$$
(reflexivity, $\mathcal{R}$)

$$\frac{X \twoheadrightarrow Y}{XU \twoheadrightarrow YV}\ V \subseteq U$$
(augmentation, $\mathcal{A}$)

$$\frac{X \twoheadrightarrow Y, Y \twoheadrightarrow Z}{X \twoheadrightarrow Z - Y}$$
~~(pseudo-transitivity, $\mathcal{T}$)~~

$$\frac{}{X \twoheadrightarrow A}\ A \in X$$
(membership, $\mathcal{M}$)

$$\frac{X \twoheadrightarrow Y}{X \twoheadrightarrow R - Y}$$
(R-complementation, $\mathcal{C}_R$)

$$\frac{}{\emptyset \twoheadrightarrow R}$$
(R-axiom, $\mathcal{C}.1$)

$$\frac{X \twoheadrightarrow Y, X \twoheadrightarrow Z}{X \twoheadrightarrow YZ}$$
(union, $\mathcal{U}$)

$$\frac{X \twoheadrightarrow Y, X \twoheadrightarrow Z}{X \twoheadrightarrow Z - Y}$$
(difference, $\mathcal{D}$)

$$\frac{X \twoheadrightarrow Y, X \twoheadrightarrow Z}{X \twoheadrightarrow Y \cap Z}$$
(intersection, $\mathcal{I}$)

- Lien: $\langle \mathcal{R}, \mathcal{A}, \mathcal{C}_R \rangle$ and exactly one of $\{\mathcal{U}, \mathcal{D}, \mathcal{I}\}$
- Objective: Is there any axiomatisation in which $\mathcal{C}_R$ can be reduced?
- Objective: How far can one reduce $\mathcal{C}_R$?

A new axiomatisation for NMVDs

Theorem 1. *The following inference rules*

$$\frac{}{\emptyset \twoheadrightarrow R} \quad (R\text{-}axiom, \mathcal{C}.1) \qquad \frac{}{A \twoheadrightarrow A} \quad (attribute\text{-}axiom, \mathcal{A}t)$$

$$\frac{X \twoheadrightarrow Y}{XA \twoheadrightarrow Y} \quad (weak\ augmentation\ rule, \mathcal{W}) \qquad \frac{X \twoheadrightarrow Y, X \twoheadrightarrow Z}{X \twoheadrightarrow Z - Y} \quad (difference\ rule, \mathcal{D})$$

form a minimal, sound and complete set of inference rules for NMVD implication.

An incomplete System

- $\mathfrak{S} = \langle \mathcal{C}.1, \mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{I} \rangle$ is incomplete
- $R = \{A, B\}$, $\Sigma = \emptyset$ and $\sigma = A \twoheadrightarrow B$
- $\sigma \in \Sigma^+_{\mathfrak{S} \cup \{\mathcal{C}\}}$
- $\sigma \notin \Sigma^+_{\mathfrak{S}}$:

	$\emptyset$	A	B	AB
$\emptyset$	$\times$			$\times$
A	$\times$	$\times$		$\times$
B	$\times$		$\times$	$\times$
AB	$\times$	$\times$	$\times$	$\times$

Strong Minimality for NMVDs

- sound and complete $\mathfrak{S}$ is *minimal* if for all $\mathfrak{R} \in \mathfrak{S}$ there is some R and some $\Sigma \cup \{\varphi\}$ on R such that $\varphi \in \Sigma^+_{\mathfrak{S}}$ but $\varphi \notin \Sigma^+_{\mathfrak{S}-\mathfrak{R}}$
- sound and complete $\mathfrak{S}$ is *strongly minimal* if for all $\mathfrak{R} \in \mathfrak{S}$ there is some R and some (trivial) φ on R such that $\varphi \in \Sigma^+_{\mathfrak{S}}$ but $\varphi \notin \Sigma^+_{\mathfrak{S}-\mathfrak{R}}$ for $\Sigma = \emptyset$

Theorem 2. *None of the minimal systems* $\langle \mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{C}_R \rangle$, $\langle \mathcal{R}, \mathcal{A}, \mathcal{I}, \mathcal{C}_R \rangle$ *and* $\langle \mathcal{R}, \mathcal{A}, \mathcal{D}, \mathcal{C}_R \rangle$ *is strongly minimal.*

Theorem 3. *The system* $\langle \mathcal{C}.1, \mathcal{A}t, \mathcal{W}, \mathcal{D} \rangle$ *is strongly minimal for NMVD implication.*

Strong Minimality for MVDs

$$\frac{}{X \twoheadrightarrow Y}\; Y \subseteq X$$
(reflexivity, $\mathcal{R}$)

$$\frac{X \twoheadrightarrow Y}{XU \twoheadrightarrow YV}\; V \subseteq U$$
(augmentation, $\mathcal{A}$)

$$\frac{X \twoheadrightarrow Y, Y \twoheadrightarrow Z}{X \twoheadrightarrow Z - Y}$$
(pseudo-transitivity, $\mathcal{T}$)

$$\frac{}{X \twoheadrightarrow A}\; A \in X$$
(membership, $\mathcal{M}$)

$$\frac{X \twoheadrightarrow Y}{X \twoheadrightarrow R - Y}$$
(R-complementation, $\mathcal{C}_R$)

$$\frac{}{\emptyset \twoheadrightarrow R}$$
(R-axiom, $\mathcal{C}.1$)

$$\frac{X \twoheadrightarrow Y, X \twoheadrightarrow Z}{X \twoheadrightarrow YZ}$$
(union, $\mathcal{U}$)

$$\frac{X \twoheadrightarrow Y, X \twoheadrightarrow Z}{X \twoheadrightarrow Z - Y}$$
(difference, $\mathcal{D}$)

$$\frac{X \twoheadrightarrow Y, X \twoheadrightarrow Z}{X \twoheadrightarrow Y \cap Z}$$
(intersection, $\mathcal{I}$)

- Mendelzon: $R\mathfrak{M} = \langle \mathcal{R}, \mathcal{T}, \mathcal{C}_R \rangle$ NOT strongly minimal
- Biskup: $R\mathfrak{B} = \langle \mathcal{C}.1, \mathcal{A}, \mathcal{T} \rangle$ strongly minimal
- Hartmann/Link: $\langle \mathcal{C}.1, \mathcal{M}, \mathcal{T} \rangle$ and exactly one of $\{\mathcal{U}, \mathcal{D}, \mathcal{I}\}$ are all strongly minimal

Conclusion

- provided a new axiomatisation for NMVDs in which 3 of the 4 inference rules have been very weakened compared to original axiomatisation
- clarifies the role of the R-complementation rule for NMVDs and extends a classical result by Biskup from total to partial relations
- introduced a stronger notion of minimality, classifies minimal axiomatisations for (N)MVDs further
- notion maybe be applied to other constraints
- MVDs for XML?
- synthesis algorithm for (N)MVDs?

Literature

- *Beeri, Fagin, Howard: A Complete Axiomatization for Functional and Multivalued Dependencies in Database Relations, SIGMOD, pp. 47-61, 1977*
- *Biskup: On the complementation rule for multivalued dependencies in database relations, Act.Inf.:10(3), pp. 297-305, 1978.*
- *Mendelzon: On axiomatising multivalued dependencies in relational databases, J. ACM:26(1), pp. 37-44, 1979.*
- *Biskup: Inferences of multivalued dependencies in fixed and undetermined universes, Theor.Comp.Sci:10(1), pp. 93-106, 1980.*
- *Lien: On the equivalence of data models, J. ACM:29(2), pp.333-363, 1982.*
- *Link: On multivalued dependencies in fixed and undetermined universes, LNCS 3861, pp. 258-277, 2006.*
- *Link: On the Logical Implication of Multivalued Dependencies with Null Values, CRPIT 51, pp. 113-122, 2006.*
- *Hartmann, Link: On a Problem of Fagin concerning Multivalued Dependencies in Relational Databases, Theor.Comp.Sci.:353(1-3), pp. 53-62, 2006*