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 \rightarrow Actor, Role

<table>
<thead>
<tr>
<th>Title</th>
<th>Actor</th>
<th>Role</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Godfather</td>
<td>Marlon Brando</td>
<td>Don Vito Corleone</td>
<td>Deleted Scene</td>
</tr>
<tr>
<td>The Godfather</td>
<td>Al Pacino</td>
<td>Michael Corleone</td>
<td>Making of</td>
</tr>
<tr>
<td>The Godfather</td>
<td>Marlon Brando</td>
<td>Don Vito Corleone</td>
<td>Making of</td>
</tr>
<tr>
<td>The Godfather</td>
<td>Al Pacino</td>
<td>Michael Corleone</td>
<td>Deleted Scene</td>
</tr>
<tr>
<td>Goldmember</td>
<td>Mike Myers</td>
<td>Austin Powers</td>
<td>Picture Gallery</td>
</tr>
<tr>
<td>Goldmember</td>
<td>Mike Myers</td>
<td>Dr. Evil</td>
<td>Subtitle</td>
</tr>
<tr>
<td>Goldmember</td>
<td>Mike Myers</td>
<td>Fat Bastard</td>
<td>Music Video</td>
</tr>
<tr>
<td>Goldmember</td>
<td>Mike Myers</td>
<td>Austin Powers</td>
<td>Subtitle</td>
</tr>
<tr>
<td>Goldmember</td>
<td>Mike Myers</td>
<td>Dr. Evil</td>
<td>Music Video</td>
</tr>
<tr>
<td>Goldmember</td>
<td>Mike Myers</td>
<td>Fat Bastard</td>
<td>Picture Gallery</td>
</tr>
<tr>
<td>Goldmember</td>
<td>Mike Myers</td>
<td>Austin Powers</td>
<td>Music Video</td>
</tr>
<tr>
<td>Goldmember</td>
<td>Mike Myers</td>
<td>Dr. Evil</td>
<td>Picture Gallery</td>
</tr>
<tr>
<td>Goldmember</td>
<td>Mike Myers</td>
<td>Fat Bastard</td>
<td>Subtitle</td>
</tr>
</tbody>
</table>

- \( r \) satisfies \( X \rightarrow Y \) if and only if \( r = r[XY] \odot r[X(R - Y)] \)

- \( \Sigma R \)-implies \( \varphi \) iff \( \forall r \subseteq dom(R) \) if \( r \) satisfies all \( \sigma \in \Sigma \), then also \( \varphi \)
Decomposition of Example

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

<table>
<thead>
<tr>
<th>Title</th>
<th>Actor</th>
<th>Role</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Godfather</td>
<td>Marlon Brando</td>
<td>Don Vito Corleone</td>
<td>Deleted Scene</td>
</tr>
<tr>
<td>The Godfather</td>
<td>Al Pacino</td>
<td>Michael Corleone</td>
<td>Making of</td>
</tr>
<tr>
<td>Goldmember</td>
<td>Mike Myers</td>
<td>Austin Powers</td>
<td>Picture Gallery</td>
</tr>
<tr>
<td>Goldmember</td>
<td>Mike Myers</td>
<td>Dr. Evil</td>
<td>Subtitle</td>
</tr>
<tr>
<td>Goldmember</td>
<td>Mike Myers</td>
<td>Fat Bastard</td>
<td>Music Video</td>
</tr>
</tbody>
</table>
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 $t_2$

• $\text{WORK}=\{\text{Employee, Child, Salary, Year}\}$ with $\text{Employee} \rightarrow \text{Child}$

<table>
<thead>
<tr>
<th>Employee</th>
<th>Child</th>
<th>Salary</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nu$</td>
<td>Maggie</td>
<td>3000</td>
<td>2005</td>
</tr>
<tr>
<td>$\nu$</td>
<td>Lisa</td>
<td>3300</td>
<td>2006</td>
</tr>
<tr>
<td>Homer</td>
<td>Bart</td>
<td>2000</td>
<td>2005</td>
</tr>
<tr>
<td>Homer</td>
<td>Lisa</td>
<td>2200</td>
<td>2006</td>
</tr>
<tr>
<td>Homer</td>
<td>Bart</td>
<td>2200</td>
<td>2006</td>
</tr>
<tr>
<td>Homer</td>
<td>Lisa</td>
<td>2000</td>
<td>2005</td>
</tr>
<tr>
<td>Mr Burns</td>
<td>$\nu$</td>
<td>8000</td>
<td>2005</td>
</tr>
<tr>
<td>Mr Burns</td>
<td>$\nu$</td>
<td>9000</td>
<td>2006</td>
</tr>
</tbody>
</table>

• $r$ satisfies $X \rightarrow Y$ 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 → Salary, Year separates \{Employee, Child, Salary, Year\} into \{Employee, Child\} and \{Employee, Salary, Year\}

<table>
<thead>
<tr>
<th>Employee</th>
<th>Child</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homer</td>
<td>Bart</td>
</tr>
<tr>
<td>Homer</td>
<td>Lisa</td>
</tr>
<tr>
<td>Mr Burns</td>
<td>(\nu)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Employee</th>
<th>Salary</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homer</td>
<td>2000</td>
<td>2005</td>
</tr>
<tr>
<td>Homer</td>
<td>2200</td>
<td>2006</td>
</tr>
<tr>
<td>Mr Burns</td>
<td>8000</td>
<td>2005</td>
</tr>
<tr>
<td>Mr Burns</td>
<td>9000</td>
<td>2006</td>
</tr>
</tbody>
</table>
**Inference Rules for MVDs**

\[
\begin{align*}
&X \rightarrow Y \\
&Y \subseteq X
\end{align*}
\]
(reflexivity, \(\mathcal{R}\))

\[
\begin{align*}
&X \rightarrow Y \\
&X U \rightarrow Y V
\end{align*}
\]
(augmentation, \(\mathcal{A}\))

\[
\begin{align*}
&X \rightarrow Y, Y \rightarrow Z \\
&X \rightarrow Z - Y
\end{align*}
\]
(pseudo-transitivity, \(\mathcal{T}\))

\[
\begin{align*}
&X \rightarrow A \\
&A \in X
\end{align*}
\]
(membership, \(\mathcal{M}\))

\[
\begin{align*}
&X \rightarrow Y \\
&X \rightarrow R - Y
\end{align*}
\]
\((R\text{-complementation, } \mathcal{C}_R)\)

\[
\begin{align*}
&\emptyset \rightarrow R
\end{align*}
\]
\((R\text{-axiom, } \mathcal{C}.1)\)

\[
\begin{align*}
&X \rightarrow Y, X \rightarrow Z \\
&X \rightarrow Y Z
\end{align*}
\]
(union, \(\mathcal{U}\))

\[
\begin{align*}
&X \rightarrow Y, X \rightarrow Z \\
&X \rightarrow Z - Y
\end{align*}
\]
(difference, \(\mathcal{D}\))

\[
\begin{align*}
&X \rightarrow Y, X \rightarrow Z \\
&X \rightarrow Y \cap Z
\end{align*}
\]
(intersection, \(\mathcal{I}\))

- Beeri, Fagin, Howard: \(R\mathcal{S} = \langle \mathcal{R}, \mathcal{A}, \mathcal{T}, \mathcal{C}_R \rangle\)
- Mendelzon: \(R\mathcal{M} = \langle \mathcal{R}, \mathcal{T}, \mathcal{C}_R \rangle\) and Biskup: \(R\mathcal{B} = \langle \mathcal{C}.1, \mathcal{A}, \mathcal{T} \rangle\)
- 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

\[
\begin{align*}
X \rightarrow Y & \quad Y \subseteq X \\
& \text{(reflexivity, } \, R) \\
XU \rightarrow YV & \quad V \subseteq U \\
& \text{(augmentation, } \, A) \\
X \rightarrow Y, \, Y \rightarrow Z & \quad X \rightarrow Z - Y \\
& \text{(pseudo transitivity, } \, T) \\
X \rightarrow A & \quad A \in X \\
& \text{(membership, } \, M) \\
X \rightarrow R - Y & \quad \emptyset \rightarrow R \\
& \text{(}R\text{-complementation, } \, C_R \text{)} \\
X \rightarrow Y, \, X \rightarrow Z & \quad X \rightarrow Y \cap Z \\
& \text{(union, } \, U) \\
X \rightarrow Y, \, X \rightarrow Z & \quad X \rightarrow Z - Y \\
& \text{(difference, } \, D) \\
X \rightarrow Y, \, X \rightarrow Z & \quad X \rightarrow Y \cap Z \\
& \text{(intersection, } \, I)
\end{align*}
\]

- Lien: \( \langle R, \, A, \, C_R \rangle \) and exactly one of \( \{U, \, D, \, I\} \)
- Objective: Is there any axiomatisation in which \( C_R \) can be reduced?
- Objective: How far can one reduce \( C_R \)?
A new axiomatisation for NMVDs

Theorem 1. The following inference rules

\[
\begin{align*}
\emptyset & \rightarrow R \\
(R\text{-axiom, } C.1) \\
X & \rightarrow Y \\
XA & \rightarrow Y \\
\text{(weak augmentation rule, } W) \\
A & \rightarrow A \\
\text{(attribute-axiom, } At) \\
X & \rightarrow Y, X \rightarrow Z \\
X & \rightarrow Z - Y \\
\text{(difference rule, } D) \\
\end{align*}
\]

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

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

<table>
<thead>
<tr>
<th></th>
<th>$\emptyset$</th>
<th>$A$</th>
<th>$B$</th>
<th>$AB$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\emptyset$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
<td></td>
</tr>
<tr>
<td>$A$</td>
<td>$\times\times$</td>
<td>$\times$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B$</td>
<td>$\times$</td>
<td>$\times\times$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$AB$</td>
<td>$\times\times\times\times$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Strong Minimality for NMVDs

- sound and complete $\mathcal{G}$ is *minimal* if for all $\mathcal{R} \in \mathcal{G}$ there is some $R$ and some $\Sigma \cup \{\varphi\}$ on $R$ such that $\varphi \in \Sigma^+_\mathcal{G}$ but $\varphi \notin \Sigma^+_{\mathcal{G}-\mathcal{R}}$

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

\[
\begin{align*}
X \rightarrow Y & \quad Y \subseteq X \quad \text{(reflexivity, } \mathcal{R}) \\
X \rightarrow Y & \quad XU \rightarrow YV \quad V \subseteq U \quad \text{(augmentation, } \mathcal{A}) \\
X \rightarrow Y & \quad X \rightarrow Z - Y \quad \text{(pseudo-transitivity, } \mathcal{T}) \\
X \rightarrow A & \quad A \in X \quad \text{(membership, } \mathcal{M}) \\
X \rightarrow R - Y & \quad X \rightarrow R - Y \quad \text{(R-complementation, } C_R) \\
\emptyset \rightarrow R & \quad \emptyset \rightarrow R \quad \text{(R-axiom, } C.1) \\
X \rightarrow Y, X \rightarrow Z & \quad X \rightarrow YZ \quad \text{(union, } \mathcal{U}) \\
X \rightarrow Y, X \rightarrow Z & \quad X \rightarrow Z - Y \quad \text{(difference, } \mathcal{D}) \\
X \rightarrow Y, X \rightarrow Z & \quad X \rightarrow Y \cap Z \quad \text{(intersection, } \mathcal{I})
\end{align*}
\]

- Mendelzon: \( R\mathcal{M} = \langle \mathcal{R}, \mathcal{T}, C_R \rangle \) NOT strongly minimal
- Biskup: \( R\mathcal{B} = \langle C.1, \mathcal{A}, \mathcal{T} \rangle \) strongly minimal
- Hartmann/Link: \( \langle C.1, \mathcal{M}, \mathcal{T} \rangle \) 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


