On Inferences of Full Hierarchical Dependencies

Sven Hartmann, Sebastian Link

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

This research is supported by the Marsden Fund Council from Government funding, administered by the Royal Society of New Zealand.
Relational Databases

- schema: finite set $R$ of attributes (with domains), e.g.
  \[\text{WORK} = \{\text{Employee, Child, Salary, Year, Insurance}\}\]

- database: $R$-relation $r$, e.g.

<table>
<thead>
<tr>
<th>Employee</th>
<th>Child</th>
<th>Salary</th>
<th>Year</th>
<th>Insurance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homer</td>
<td>Bart</td>
<td>2000</td>
<td>2006</td>
<td>National</td>
</tr>
<tr>
<td>Homer</td>
<td>Lisa</td>
<td>2200</td>
<td>2007</td>
<td>National</td>
</tr>
<tr>
<td>Homer</td>
<td>Bart</td>
<td>2200</td>
<td>2006</td>
<td>National</td>
</tr>
<tr>
<td>Homer</td>
<td>Lisa</td>
<td>2000</td>
<td>2007</td>
<td>National</td>
</tr>
<tr>
<td>Homer</td>
<td>Bart</td>
<td>2000</td>
<td>2006</td>
<td>AMI</td>
</tr>
<tr>
<td>Homer</td>
<td>Lisa</td>
<td>2200</td>
<td>2007</td>
<td>AMI</td>
</tr>
<tr>
<td>Homer</td>
<td>Bart</td>
<td>2200</td>
<td>2006</td>
<td>AMI</td>
</tr>
<tr>
<td>Homer</td>
<td>Lisa</td>
<td>2000</td>
<td>2007</td>
<td>AMI</td>
</tr>
</tbody>
</table>

- attribute $A$ with $\text{dom}(A)$, attribute sets: $W, X, Y, Z, Y_i \ (\subseteq R)$
- $r[X], \text{Dom}(r)$
Full Hierarchical Dependencies

- FHDs $X : \{Y_1, \ldots, Y_k\}$ with mutually disjoint $X, Y_1, \ldots, Y_k, k \geq 1$
  for example, Employee: $\{\{\text{Child}\}, \{\text{Insurance}\}\}$

- $\models_r X : \{Y_1, \ldots, Y_k\}$ precisely when $r$ is decomposable into $k + 1$ of
  its projections without loss of information

$$r = r[XY_1] \bowtie \cdots \bowtie r[XY_k] \bowtie r[X(R - Y_1 \cdots Y_k)]$$

- Employee: $\{\{\text{Child}\}, \{\text{Insurance}\}\}$ satisfied by previous database

- store independent facts separately (less redundancy, better updating)

<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>
</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>2006</td>
</tr>
<tr>
<td>Homer</td>
<td>2200</td>
<td>2007</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Employee</th>
<th>Insurance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homer</td>
<td>National</td>
</tr>
<tr>
<td>Homer</td>
<td>AMI</td>
</tr>
</tbody>
</table>
Hierarchical Decompositions

- FHDs permit stepwise decompositions dividing each component into two new ones
- apply Employee: \{\{Child\},\{Insurance\}\} to \{Employee,Child,Salary,Year,Insurance\}:

\[
\begin{array}{c}
\text{Employee, Child, Insurance, Salary, Year} \\
\text{Employee, Child} \\
\text{Employee, Insurance, Salary, Year} \\
\text{Employee, Child, Insurance, Salary, Year} \\
\text{Employee, Insurance} \\
\text{Employee, Child, Salary, Year} \\
\text{Employee, Child, Insurance, Salary, Year} \\
\text{Employee, Child} \\
\text{Employee, Salary, Year} \\
\text{Employee, Child, Insurance, Salary, Year} \\
\text{Employee, Child} \\
\text{Employee, Salary, Year}
\end{array}
\]

- join dependencies do not share this feature
Axiomatisation for FHDs in Fixed Universes

- $\Sigma \cup \{X : \{Y_1, \ldots, Y_k\}\}$ set of FHDs on $R$ (all attributes in $R$)
  $\Sigma$ $R$-implies $X : \{Y_1, \ldots, Y_k\}$ iff $\forall r \subseteq dom(R)$: if $\models_r \Sigma$, then $\models_r X : \{Y_1, \ldots, Y_k\}$

  $\emptyset : \{\emptyset\}$
  (empty-set-axiom, $R_\emptyset$)

  $X : \{Y_1, \ldots, Y_k, Y\}$
  $X : \{Y_1, \ldots, Y_k\}$
  (omission, $O$)

  $X : \{Y_1, \ldots, Y_k\}$
  $X Z : \{Y_1 - Z, \ldots, Y_k - Z\}$
  (augmentation, $A$)

  $X : \{Y_1, \ldots, Y_k\}$
  $X : \{Y_1, \ldots, Y_k, Y\}$
  (transitivity, $T$)

  $\emptyset : \{R\}$
  (R-axiom)

  $X : \{Y_1, \ldots, Y_k\}$
  $X : \{Y_1, \ldots, Y_k\}$
  (R-complementation, $C_R$)

  $X : \{Y_1, \ldots, Y_k\}$
  $X : \{Y_1, \ldots, Y_k, Y\}$
  (R-complementation, $C_R$)

  $\emptyset : \{R\}$
  (R-axiom)

  $X : \{Y_1, \ldots, Y_k\}$
  $X : \{Z\}$
  (union, $\cup$)

  $X : \{Y_1, \ldots, Y_k\}$
  $X : \{Y_1, \ldots, Y_k, Y\}$
  (union, $\cup$)

  $X : \{Y_1, \ldots, Y_k\}$
  $X : \{Z\}$
  (difference, $\setminus$)

  $X : \{Y_1, \ldots, Y_k\}$
  $X : \{Y_1, \ldots, Y_k\}$
  (difference, $\setminus$)

  $X : \{Y_1, \ldots, Y_k\}$
  $X : \{Z\}$
  (intersection, $\cap$)

  $X : \{Y_1, \ldots, Y_k\}$
  $X : \{Z\}$
  (intersection, $\cap$)

- Theorem:
  $\mathcal{H} = \langle R_\emptyset, A, T, O, C_R \rangle$ $R$-sound & $R$-complete ($\forall \Sigma$ on $R$: $\Sigma^+_\mathcal{H} = \Sigma^*_R$)
The Role of the Complementation Rule

- complementation rule $C_R$ enjoys special status
- $\text{Employee:}\{\{\text{Child}\},\{\text{Insurance}\}\} \models \text{Employee:}\{\{\text{Child}\},\{\text{Salary, Year}\}\}$?
- the answer depends on the underlying relation schema $R$

  - on $R = \{\text{Employee, Child, Insurance, Salary, Year}\}$: yes!
  - on $R = \{\text{Employee, Child, Insurance, Salary, Year, Office}\}$: no!

- $C_R$ just a means of database normalisation on universe $R$

  - two ways to go:
    - $\leftrightarrow$ the status of $C_R$ should be reflected within axiomatisation
    - $\leftrightarrow$ find notion of implication independent from underlying schema
Complementarity

- $R$-complementary $\mathcal{C}$ apply $C_R$ at most in very last inference step
- $\Sigma = \{ \text{Employee:}\{\{\text{Child}\},\{\text{Insurance}\}\}, \text{Employee:}\{\{\text{Salary, Year}\}\} \}$
- Employee:{$\{\text{Child}\}, \{\text{Insurance, Salary, Year}\}$} $\notin \Sigma^+_{\{R_\emptyset, A, O, T\}}$
- Employee:{$\{\text{Child}\}, Y\} \notin \Sigma^+_{\{R_\emptyset, A, O, T\}}$
  \[ \forall Y. Y \in \{ \text{Employee, Child, Insurance, Salary, Year} \} \neq \emptyset \]
- for $R := \{ \text{Employee, Child, Insurance, Salary, Year, Office} \}$ we have
  Employee:{$\{\{\text{Child}\}, \{\text{Insurance, Salary, Year}\}\}$} $\in \Sigma^+_{\mathcal{C}}$
- in any such inference $C_R$ must be used at least once, but $R = \{ \text{Employee, Insurance, Child, Salary, Year} \} = \{ \text{Office} \}$ implies that $C_R$ is not just used as last rule
Example Derivation

- using $\langle R_{\emptyset}, A, O, T, C_R \rangle$:
  infer the FHD Employee:{$\{\text{Child}\}, \{\text{Insurance,Salary,Year}\}$} from Employee:{$\{\text{Child}\}, \{\text{Insurance}\}$} & Employee:{$\{\text{Salary,Year}\}$}

Employee : {\{Child\}, \{Insurance\}}

---

Employee : {\{Child\}, \{Salary,Year,Office\}}

---

Employee,Salary,Year : {\{Child\}, \{Office\}}

---

Employee : {\{Child\}, \{Office\}, \{Salary,Year\}}

---

Employee : {\{Child\}, \{Office\}}

---

Employee : {\{Child\}, \{Insurance,Salary,Year\}}

---

the minimal system $\mathcal{H} = \langle R_{\emptyset}, A, O, T, C_R \rangle$ is not complementary
Complementary Axiomatisation for FHDs

- require two additional inference rules:

\[
\begin{align*}
X : \{Y\}, \quad W : \{Y_1, \ldots, Y_k\} \\
\therefore Y \cap W = \emptyset \\
\quad X : \{Y \cap Y_1, \ldots, Y \cap Y_k, Y - Y_1 \cdots Y_k\} \\
\quad (\text{subset rule, } \mathcal{S})
\end{align*}
\]

\[
\begin{align*}
X : \{Y_1, \ldots, Y_k, Y_{k+1}\} \\
\therefore X : \{Y_1, \ldots, Y_k Y_{k+1}\} \\
\quad (\text{merging rule, } \mathcal{M})
\end{align*}
\]

- **Theorem:**
  \( \mathcal{H}_C = \langle R_\emptyset, A, T, O, S, M, C_R \rangle \) is \( R \)-complete and \( R \)-complementary for the \( R \)-implication of FHDs for all \( R \)

- **Theorem:**
  \( \mathcal{H}_C \) minimal in the sense that the omission of any of its rules leads to the loss of \( R \)-completeness or \( R \)-complementarity for at least some \( R \)

- subsystem \( \mathcal{H}_U = \langle R_\emptyset, A, T, O, S, M \rangle \) is nearly \( R \)-complete
Back to our Example Derivation

- using $\mathcal{H}_C = \langle R_\emptyset, A, O, T, S, M \rangle$:
  infer the FHD $\text{Employee:}\{\{\text{Child}\}, \{\text{Insurance,Salary,Year}\}\}$ from
  $\text{Employee:}\{\{\text{Child}\}, \{\text{Insurance}\}\}$ & $\text{Employee:}\{\{\text{Salary,Year}\}\}$

\[
\begin{align*}
\text{Employee : } \{\{\text{Child}\}, \{\text{Insurance}\}\} \\
\text{Employee,Salary,Year : } \{\{\text{Child}\}, \{\text{Insurance}\}\}^A \\
\text{Employee : } \{\{\text{Salary,Year}\}\}^T \\
\text{Employee : } \{\{\text{Child}\}, \{\text{Insurance,Salary,Year}\}\}^M
\end{align*}
\]

- in using $\mathcal{H}_C$ universes can be fixed at the very last step of the inference
Another Inference

- using \( \mathcal{H} \) and then \( \mathcal{H}_C \), respectively:
  infer the FHD Employee: \{|\{\text{Child}\}, \{\text{Salary,Year}\}\} \) from 
  Employee: \{|\{\text{Insurance,Salary,Year}\}\} \& \text{Employee,Child:}\{|\{\text{Salary,Year}\}\}\)
Implication in undetermined Universes

- consequences dependent on the universe are in fact no consequences

- expr $X : \{Y_1, \ldots, Y_k\}$ with finite mutually disjoint $X, Y_1, \ldots, Y_k$

- $\models_r X : \{Y_1, \ldots, Y_k\}$ iff $X \cup Y_1 \cup \cdots Y_k \subseteq \text{Dom}(r)$ and

  $$r = r[XY_1] \bowtie \cdots \bowtie r[XY_k] \bowtie r[X \cup \text{Dom}(r) - Y_1 \cdots Y_k]$$

- $\Sigma \models \varphi$ iff for every $r$ to which all FHDs in $\Sigma \cup \\{\varphi\}$ can be applied we have: if $\models_r \Sigma$, then $\models_r \varphi$

- if $R$ contains all attributes occurring in any FHD from $\Sigma \cup \\{\varphi\}$: $\Sigma$ $R$-implies $\varphi$ whenever $\Sigma$ implies $\varphi$, but not vice versa!
Axiomatising FHDs in undetermined Universes

- Employee:{\{Child\},\{Insurance\}} $R$-implies Employee:{\{Child\},\{Salary,Year\}} for $R = \{\text{Employee,Child,Insurance,Salary,Year}\}$

- Employee:{\{Child\},\{Insurance\}} $\not\models$ Employee:{\{Child\},\{Salary,Year\}}:

<table>
<thead>
<tr>
<th>Employee</th>
<th>Child</th>
<th>Insurance</th>
<th>Salary</th>
<th>Year</th>
<th>Office</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kamini</td>
<td>Casandra</td>
<td>Asteron</td>
<td>35</td>
<td>2006</td>
<td>01.09</td>
</tr>
<tr>
<td>Kamini</td>
<td>Casandra</td>
<td>Asteron</td>
<td>38</td>
<td>2007</td>
<td>02.09</td>
</tr>
</tbody>
</table>

- **Theorem**: $\mathcal{H}_U = \langle \mathcal{R}_\emptyset, \mathcal{A}, \mathcal{T}, \mathcal{O}, \mathcal{S}, \mathcal{M} \rangle$ is sound and complete for the implication of FHDs in undetermined universes

- **Theorem**: no proper subset of $\mathcal{H}_U = \langle \mathcal{R}_\emptyset, \mathcal{A}, \mathcal{T}, \mathcal{O}, \mathcal{S}, \mathcal{M} \rangle$ is complete

- **notice**: $\mathcal{H}_C = \mathcal{H}_U \cup \{\mathcal{C}_R\}$
Interesting Things to look at

- Are there any minimal sets of inference rules that are also complementary?

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

- Consider complete axiomatisations of MVDs in Nested Database Models. Are these complementary?

- Consider complete axiomatisations of fuzzy and approximate MVDs. Are these complementary?

- synthesis algorithms for MVDs and FHDs?

- views?