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.

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

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

Employee	Child	Salary	Year	Insurance
Homer	Bart	2000	2006	National
Homer	Lisa	2200	2007	National
Homer	Bart	2200	2006	National
Homer	Lisa	2000	2007	National
Homer	Bart	2000	2006	AMI
Homer	Lisa	2200	2007	AMI
Homer	Bart	2200	2006	AMI
Homer	Lisa	2000	2007	AMI

- attribute A with $dom(A)$, attribute sets: W, X, Y, Z, Y_i $(\subseteq R)$
- $r[X]$, $Dom(r)$

MARSDEN FUND
TE PŪTEA RANGAHAU A MARSDEN

Full Hierarchical Dependencies

- FHDs $X : \{Y_1, \dots, Y_k\}$ with mutually disjoint $X, Y_1, \dots, Y_k, k \geq 1$ for example, Employee: {{Child},{Insurance}}
- $\models_r X : \{Y_1, \dots, Y_k\}$ 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:{{Child},{Insurance}} satisfied by previous database
- store independent facts separately (less redundancy, better updating)

Employee	Child
Homer	Bart
Homer	Lisa

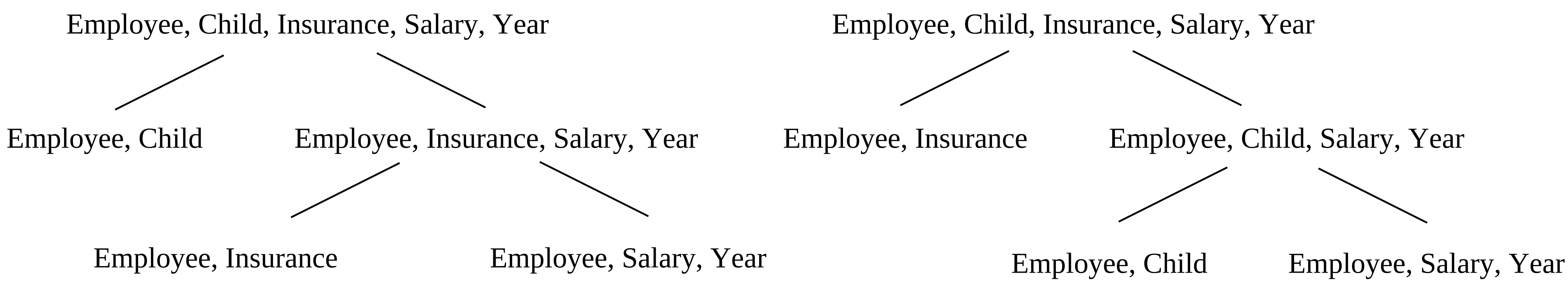
Employee	Salary	Year
Homer	2000	2006
Homer	2200	2007

Employee	Insurance
Homer	National
Homer	AMI

Hierarchical Decompositions

- FHDs permit stepwise decompositions dividing each component into two new ones

- apply Employee:{{Child},{Insurance}} to {Employee,Child,Salary,Year,Insurance}:



- 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)
  Σ R-implies $X : \{Y_1, \ldots, Y_k\}$ iff $\forall r \subseteq dom(R)$: if $\models_r \Sigma$, then $\models_r X : \{Y_1, \ldots, Y_k\}$

$$\frac{}{\emptyset : \{\emptyset\}} \quad \text{(empty-set-axiom, } \mathcal{R}_\emptyset\text{)}$$

$$\frac{X : \{Y_1, \ldots, Y_k\}}{XZ : \{Y_1 - Z, \ldots, Y_k - Z\}} \quad \text{(augmentation, } \mathcal{A}\text{)}$$

$$\frac{XY : \{Y_1, \ldots, Y_k\}, X : \{Y\}}{X : \{Y_1, \ldots, Y_k, Y\}} \quad \text{(transitivity, } \mathcal{T}\text{)}$$

$$\frac{X : \{Y_1, \ldots, Y_k, Y\}}{X : \{Y_1, \ldots, Y_k\}} \quad \text{(omission, } \mathcal{O}\text{)}$$

$$\frac{X : \{Y_1, \ldots, Y_k\}}{X : \{Y_1, \ldots, Y_{k-1}, R - XY_1 \cdots Y_k\}} \quad \text{(}R\text{-complementation, } \mathcal{C}_R\text{)}$$

$$\frac{}{\emptyset : \{R\}} \quad \text{(}R\text{-axiom)}$$

$$\frac{X : \{Y_1, \ldots, Y_k\}, X : \{Z\}}{X : \{Y_1 - Z, \ldots, Y_{k-1} - Z, Y_k Z\}} \quad \text{(union, } \mathcal{U}\text{)}$$

$$\frac{X : \{Y_1, \ldots, Y_k\}, X : \{Z\}}{X : \{Y_1, \ldots, Y_{k-1}, Y_k - Z\}} \quad \text{(difference, } \mathcal{D}\text{)}$$

$$\frac{X : \{Y_1, \ldots, Y_k\}, X : \{Z\}}{X : \{Y_1, \ldots, Y_{k-1}, Y_k \cap Z\}} \quad \text{(intersection, } \mathcal{I}\text{)}$$

- Theorem:
  $\mathfrak{H} = \langle \mathcal{R}_\emptyset, \mathcal{A}, \mathcal{T}, \mathcal{O}, \mathcal{C}_R \rangle$ R-sound & R-complete ($\forall \Sigma$ on R: $\Sigma^+_{\mathfrak{H}} = \Sigma^*_R$)

The Role of the Complementation Rule

- complementation rule $\mathcal{C}_R$ enjoys special status
- Employee:{{Child},{Insurance}}$\models$Employee:{{Child},{Salary,Year}}?
- the answer depends on the underlying relation schema R
- on $R = \{$Employee,Child,Insurance,Salary,Year$\}$: yes!
- on $R = \{$Employee,Child,Insurance,Salary,Year,Office$\}$: no!
- $\mathcal{C}_R$ just a means of database normalisation on universe R
- two ways to go:
  - $\hookrightarrow$ the status of $\mathcal{C}_R$ should be reflected within axiomatisation
  - $\hookrightarrow$ find notion of implication independent from underlying schema

Complementarity

- *R-complementary* $\mathfrak{S}$ apply $\mathcal{C}_R$ at most in very last inference step
- Σ ={Employee:{{Child},{Insurance}},Employee:{{Salary,Year}}}
- Employee:{{Child},{Insurance,Salary,Year}}$\notin \Sigma^+_{\{\mathcal{R}_\emptyset,\mathcal{A},\mathcal{O},\mathcal{T}\}}$
- Employee:{{Child},Y}$\notin \Sigma^+_{\{\mathcal{R}_\emptyset,\mathcal{A},\mathcal{O},\mathcal{T}\}}$

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

- for R := {Employee,Child,Insurance,Salary,Year,Office} we have

$$\text{Employee:}\{\{\text{Child}\},\{\text{Insurance,Salary,Year}\}\} \in \Sigma^+_{\mathfrak{H}}$$

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

Example Derivation

- using $\langle \mathcal{R}_\emptyset, \mathcal{A}, \mathcal{O}, \mathcal{T}, \mathcal{C}_R \rangle$:
  infer the FHD Employee:{{Child},{Insurance,Salary,Year}} from Employee:{{Child},{Insurance}} & Employee:{{Salary,Year}}

$$
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\text{Employee} : \{\{\text{Child}\}, \{\text{Insurance}\}\}}{\text{Employee} : \{\{\text{Child}\}, \{\text{Salary,Year,Office}\}\}}\mathcal{C}_R}{\text{Employee,Salary,Year} : \{\{\text{Child}\}, \{\text{Office}\}\}}\mathcal{A} \quad \text{Employee} : \{\{\text{Salary,Year}\}\}}{\text{Employee} : \{\{\text{Child}\}, \{\text{Office}\}, \{\text{Salary,Year}\}\}}\mathcal{T}}{\text{Employee} : \{\{\text{Child}\}, \{\text{Office}\}\}}\mathcal{O}}{\text{Employee} : \{\{\text{Child}\}, \{\text{Insurance,Salary,Year}\}\}}\mathcal{C}_R
$$

- the minimal system $\mathfrak{H} = \langle \mathcal{R}_\emptyset, \mathcal{A}, \mathcal{O}, \mathcal{T}, \mathcal{C}_R \rangle$ is not complementary

Complementary Axiomatisation for FHDs

- require two additional inference rules:

$$\frac{X:\{Y\},\ W:\{Y_1,\ldots,Y_k\}}{X:\{Y\cap Y_1,\ldots,Y\cap Y_k,Y-Y_1\cdots Y_k\}}Y\cap W=\emptyset \qquad \frac{X:\{Y_1,\ldots,Y_k,Y_{k+1}\}}{X:\{Y_1,\ldots,Y_kY_{k+1}\}}$$

(subset rule, $\mathcal{S}$) (merging rule, $\mathcal{M}$)

- Theorem:
  $\mathfrak{H}_{\mathcal{C}} = \langle \mathcal{R}_\emptyset, \mathcal{A}, \mathcal{T}, \mathcal{O}, \mathcal{S}, \mathcal{M}, \mathcal{C}_R \rangle$ is R-complete and R-complementary for the R-implication of FHDs for all R

- Theorem:
  $\mathfrak{H}_{\mathcal{C}}$ 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 $\mathfrak{H}_{\mathcal{U}} = \langle \mathcal{R}_\emptyset, \mathcal{A}, \mathcal{T}, \mathcal{O}, \mathcal{S}, \mathcal{M} \rangle$ is nearly R-complete

Back to our Example Derivation

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

$$\dfrac{\dfrac{\text{Employee} : \{\{\text{Child}\}, \{\text{Insurance}\}\}}{\text{Employee,Salary,Year} : \{\{\text{Child}\}, \{\text{Insurance}\}\}}\mathcal{A} \qquad \text{Employee} : \{\{\text{Salary,Year}\}\}}{\dfrac{\text{Employee} : \{\{\text{Child}\}, \{\text{Insurance}\}, \{\text{Salary,Year}\}\}}{\text{Employee} : \{\{\text{Child}\}, \{\text{Insurance,Salary,Year}\}\}}\mathcal{M}}\mathcal{T}$$

- in using $\mathfrak{H}_{\mathcal{C}}$ universes can be fixed at the very last step of the inference

Another Inference

- using $\mathfrak{H}$ and then $\mathfrak{H}_{\mathcal{C}}$, respectively:
  infer the FHD Employee:{{Child},{Salary,Year}} from
  Employee:{{Insurance,Salary,Year}}&Employee,Child:{{Salary,Year}}

$$\dfrac{\dfrac{\text{Employee} : \{\{\text{Insurance,Salary,Year}\}\}}{\text{Employee} : \{\{\text{Child}\}\}}\mathcal{C}_R \quad \text{Employee,Child} : \{\{\text{Salary,Year}\}\}}{\text{Employee} : \{\{\text{Salary,Year}\}, \{\text{Child}\}\}}\mathcal{T}$$

$$\dfrac{\dfrac{\text{Employee} : \{\{\text{Insurance,Salary,Year}\}\} \quad \text{Employee,Child} : \{\{\text{Salary,Year}\}\}}{\text{Employee} : \{\{\text{Salary,Year}\}, \{\text{Insurance}\}\}}\mathcal{S}}{\text{Employee} : \{\{\text{Salary,Year}\}, \{\text{Child}\}\}}\mathcal{C}_R$$

- subset rule $\mathcal{S}$ needed to shift $\mathcal{C}_R$ to very end of inference

Implication in undetermined Universes

- consequences dependent on the universe are in fact no consequences
- expr $X : \{Y_1, \dots, Y_k\}$ with finite mutually disjoint $X, Y_1, \dots, Y_k$
- $\models_r X : \{Y_1, \dots, Y_k\}$ iff $X \cup Y_1 \cup \cdots Y_k \subseteq Dom(r)$ and

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

- $\Sigma \models \varphi$ 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 $\Sigma \cup \{\varphi\}$: Σ R-implies φ whenever Σ implies φ, but not vice versa!

Axiomatising FHDs in undetermined Universes

- Employee:{{Child},{Insurance}} R-implies Employee:{{Child},{Salary,Year}} for $R = \{$Employee,Child,Insurance,Salary,Year$\}$
- Employee:{{Child},{Insurance}} $\not\models$ Employee:{{Child},{Salary,Year}}:

Employee	Child	Insurance	Salary	Year	Office
Kamini	Casandra	Asteron	35	2006	01.09
Kamini	Casandra	Asteron	38	2007	02.09

- Theorem: $\mathfrak{H}_{\mathcal{U}} = \langle \mathcal{R}_{\emptyset}, \mathcal{A}, \mathcal{T}, \mathcal{O}, \mathcal{S}, \mathcal{M} \rangle$ is sound and complete for the implication of FHDs in undetermined universes
- Theorem: no proper subset of $\mathfrak{H}_{\mathcal{U}} = \langle \mathcal{R}_{\emptyset}, \mathcal{A}, \mathcal{T}, \mathcal{O}, \mathcal{S}, \mathcal{M} \rangle$ is complete
- notice: $\mathfrak{H}_{\mathcal{C}} = \mathfrak{H}_{\mathcal{U}} \cup \{\mathcal{C}_R\}$

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?