Functional Dependencies on Nested Attributes: Algebraic, Logical and Topological Perspective

Sven Hartmann, Sebastian Link

Information Science Research Centre,
Massey University, Palmerston North,
New Zealand
Outline

1. Background and Data Models
2. The Algebra of Nested Attributes
3. Algebraic Perspective
4. Logical Perspective
5. Topological Perspective
6. Conclusion and Future Work
1.1 Functional Dependencies in the RDM

- FDs introduced in context of RDM by E.F. Codd in 1972
  - expression $X \rightarrow Y$ with $X, Y \subseteq R$
  - $\models_r X \rightarrow Y$ iff $t_1[Y] = t_2[Y]$, if $t_1[X] = t_2[X]$ for any $t_1, t_2 \in r$
- gain complete knowledge about consequences of semantics specified
- **Boolean Algebra** $(\mathcal{P}(R), \subseteq, \cup, \cap, -, \emptyset, R)$ on $R$
- **Armstrong Axioms** (1974)

\[
\begin{align*}
Y \subseteq X & \quad X \rightarrow Y \subseteq X \\
X \rightarrow X \cup Y & \quad X \rightarrow Y, Y \rightarrow Z \\
& \quad X \rightarrow Z
\end{align*}
\]
- Fagin 1977: implication is equivalent to that of Horn clauses
- impact:
  - implication problem, equiv of sets of FDs, minimal covers
  - normal forms, redundancies, update anomalies, integrity checking
1.2 Advanced Data Models

- ER, UML, HERM, Nested RDM, Object-oriented/relational, XML
- extend achievements to complex objects in unified framework
- classify data models by the types they support
2.1 Database Schemata: Nested Attributes

- capture characteristics of objects in target database by attributes

\[ N := A | \lambda | L(N_1, \ldots, N_k) | L[N] | L\{N\} | L\langle N \rangle \]

- examples:
  - Shop(Customer, Trolley(Item(Article, Price)), Discount)
  - Soccer{Match(Winner, Loser)}
  - Enrolments[Student(ID, Name, History[Course(No, Name, Grade)]]]
2.2 Database Instances: Domain Assignment

- extend \( dom \) from flat to nested attributes \( (\text{dom}(\lambda) = \{ \text{ok} \}) \)
- examples for nested tuples:
  - \( \text{Shop} (\text{Customer}, \text{Trolley} \langle \text{Item} (\text{Article}, \text{Price}) \rangle, \text{Discount}) \):
    - \( \langle \text{Homer}, \langle (\text{Donut}, 1.5), (\text{Donut}, 1.5), (\text{Chocolate}, 2), (\text{Chocolate}, 2) \rangle, 0 \rangle \)
    - \( \langle \text{Bart}, \langle (\text{Donut}, 2), (\text{Donut}, 2), (\text{Chocolate}, 1.5), (\text{Chocolate}, 1.5) \rangle, 1 \rangle \)
  - \( \text{Soccer} \{ \text{Match} (\text{Winner}, \text{Loser}) \} \):
    - \( \{ (\text{Denmark}, \text{Sweden}), (\text{New Zealand}, \text{Australia}) \} \)
    - \( \{ (\text{Mexico}, \text{USA}), (\text{Brazil}, \text{Argentina}), (\text{Brazil}, \text{USA}) \} \)

- RDM: single application of record constructor
- Nested Relational Data Model: record and set constructor
- Object-oriented Data Models: record, set, multiset and list constructor
2.3 Subschemata: Subattributes

• recursively replacing attributes by $\lambda$ gives different layers of info:

• some subattributes of
  Shop(Customer,Trolley(Item(Article,Price)),Discount):
    • Shop($\lambda$,Trolley(Item(Article,Price)),Discount)
    • Shop(Customer,Trolley(Item($\lambda$, $\lambda$)),Discount)
    • Shop($\lambda$,Trolley(Item(Article,$\lambda$)),$\lambda$)
    • Shop(Customer,$\lambda$,Discount)

• formally:
  define subattribute relation $\leq$ on nested attributes (partial order)
### 2.4 Database Transformations: Projection Function

- Subattributes represent at most as much info as their superattributes.
- Formally: for $M \leq N$ there is projection $\pi_M^N : \text{dom}(N) \rightarrow \text{dom}(M)$.

Let $N = \text{Shop}($Customer$, \text{Trolley}(<\text{Item}(\text{Article}, \text{Price})), \text{Discount})$ with $t = (\text{Bart}, \langle (\text{Donut}, 2), (\text{Donut}, 2), (\text{Chocolate}, 1.5), (\text{Chocolate}, 1.5) \rangle, 1)$.

- $M = \text{Shop}($Customer$, \text{Trolley}(<\text{Item}(\lambda, \text{Price})), \text{Discount})$

  \[
  \pi_M^N(t) = (\text{Bart}, \langle (\text{ok}, 2), (\text{ok}, 2), (\text{ok}, 1.5), (\text{ok}, 1.5) \rangle, 1)
  \]

- $M = \text{Shop}(\lambda, \text{Trolley}(<\text{Item}(\lambda, \lambda)), \text{Discount})$

  \[
  \pi_M^N(t) = (\text{ok}, \langle (\text{ok}, \text{ok}), (\text{ok}, \text{ok}), (\text{ok}, \text{ok}), (\text{ok}, \text{ok}) \rangle, 1)
  \]
2.5 The Brouwerian Algebra of Subattributes

- subattribute order \( \leq \) induces operations \( \sqcup_N, \sqcap_N, \) and \( \neg_N \)
- \( (\text{Sub}(N) = \{M \mid M \leq N\}, \leq, \sqcup_N, \sqcap_N, \neg_N, N) \) is Brouwerian Algebra
  - \( (\text{Sub}(N), \leq, \sqcup_N, \sqcap_N) \) is a lattice
  - \( N \) is top element
  - pseudo difference \( Z \neg Y \) of \( Z \) and \( Y \) in \( \text{Sub}(N) \) satisfies
    \[
    Z \neg Y \leq X \text{ if and only if } Z \leq Y \sqcup X
    \]
    for all \( X \in \text{Sub}(N) \)
- **Brouwerian Complement**: \( Y^C_N = N \neg_N Y \) satisfies
  \[
  Y^C \leq X \text{ if and only if } X \sqcup Y = N
  \]
- \( (\text{Sub}(N), \leq, \sqcup_N, \sqcap_N, (\cdot)^C_N, \lambda_N, N) \) is not a Boolean Algebra
2.6 The Algebra of Nested Attributes: An Example
2.7 The Algebra of Nested Attributes: A further Example
3.1 Functional Dependencies

- **functional dependency** on nested attribute $N$ is
  \[ \mathcal{X} \rightarrow \mathcal{Y} \quad \text{with} \quad \mathcal{X}, \mathcal{Y} \subseteq \text{Sub}(N) \text{ non-empty} \]

- $r \subseteq \text{Dom}(N)$ satisfies $\mathcal{X} \rightarrow \mathcal{Y}$ on $N$ ($\models_r \mathcal{X} \rightarrow \mathcal{Y}$) iff
  \[ \pi_X^N(t_1) = \pi_X^N(t_2) \forall X \in \mathcal{X} \quad \text{implies} \quad \pi_Y^N(t_1) = \pi_Y^N(t_2) \forall Y \in \mathcal{Y} \]

- $\text{Shop}(\lambda, \text{Trolley} \langle \text{Item(Article,Price)} \rangle, \lambda) \rightarrow \text{Shop}(\lambda, \lambda, \text{Discount})$

- $\{ \text{Shop}(\lambda, \text{Trolley} \langle \text{Item(Article,\lambda)} \rangle, \lambda), \text{Shop}(\lambda, \text{Trolley} \langle \text{Item(\lambda,Price)} \rangle, \lambda) \} \rightarrow \text{Shop}(\lambda, \lambda, \text{Discount})$

- implication: $\Sigma \models \tau$ iff $\models_r \tau$ if $\models_r \sigma$ for all $\sigma \in \Sigma$ and any (finite) $r$

- goal: find **sound** and **complete** $\mathcal{R}$, i.e., $\Sigma_\mathcal{R}^+ \subseteq \Sigma^*$ and $\Sigma^* \subseteq \Sigma_\mathcal{R}^+$
3.2 A fundamental Difference

- $N = \text{Soccer}\{\text{Match}(\text{Winner}, \text{Loser})\}$
- $r = \{t_1, t_2\} \subseteq \text{Dom}(N)$ with
  - $t_1 = \{(\text{Denmark, Brazil}), (\text{Germany, Italy})\}$ and
  - $t_2 = \{(\text{Denmark, Italy}), (\text{Germany, Brazil})\}$
- $\models_r \text{Soccer}\{\text{Match}(\text{Winner})\} \rightarrow \text{Soccer}\{\text{Match}(\text{Loser})\}$
- $\not\models_r \text{Soccer}\{\text{Match}(\text{Winner})\} \rightarrow \text{Soccer}\{\text{Match}(\text{Winner, Loser})\}$
- values on subattributes $X$ and $Y$ do not determine values on $X \sqcup Y$
- the bad guys are: sets and multisets
- shows: FDs cannot be simplified to $X \rightarrow Y$ with $X, Y \in \text{Sub}(N)$
- FDs simpler in case of records and lists only
3.3 Reconcilable Attributes

- $X, Y \in \text{Sub}(N)$ reconcilable iff one of the following holds:
  
  - $Y \leq X$ or $X \leq Y$,
  
  - $N = L(N_1, \ldots, N_k), X = L(X_1, \ldots, X_k), Y = L(Y_1, \ldots, Y_k)$
    where $X_i$ and $Y_i$ are reconcilable for all $i = 1, \ldots, k$,
  
  - $N = L[N'], X = L[X'], Y = L[Y']$ where $X'$ and $Y'$ reconcilable
  
  - Soccer{$\text{Match}(\text{Winner}, \lambda)$}, Soccer{$\text{Match}(\lambda, \text{Loser})$} not reconcilable
  
  - Shop$(\lambda, \text{Trolley}($Item$(\text{Article}, \lambda)), \lambda), \text{Shop}(\lambda, \text{Trolley}($Item$(\lambda, \text{Price})), \lambda)$
3.4 Axiomatisation

Theorem 1. Let $N \in \mathcal{N}A$ and $X, Y, Z \in \text{Sub}(N)$. The Armstrong Axioms, i.e.,

$$X \rightarrow Y \quad Y \leq X,$$

$$X \rightarrow Y \quad X \rightarrow X \cup_N Y,$$

$$X \rightarrow Y, Y \rightarrow Z \quad X \rightarrow Z$$

form a minimal, sound and complete set of inference rules for the implication of FDs in the presence of records, and records and lists.

Let $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \subseteq \text{Sub}(N)$ be non-empty, and $\mathcal{T}$ be any non-empty subset of \{lists, sets, multisets\} apart from \{lists\}. The generalised Armstrong Axioms, i.e.,

$$\mathcal{X} \rightarrow \mathcal{Y} \quad \mathcal{Y} \subseteq \mathcal{X},$$

$$\{X\} \rightarrow \{Y\} \quad Y \leq X,$$

$$\mathcal{X} \rightarrow \mathcal{X} \cup \mathcal{Y},$$

$$\{X, Y\} \rightarrow \{X \cup_N Y\} \quad X, Y \text{ reconcilable},$$

$$\mathcal{X} \rightarrow \mathcal{Y}, \mathcal{Y} \rightarrow \mathcal{Z} \quad \mathcal{X} \rightarrow \mathcal{Z}$$

form a minimal, sound and complete set of inference rules for the implication of FDs in the presence of records and $\mathcal{T}$. □
4.1 Join-Irreducibles

- one idea behind Fagin’s Equivalence theorem: interpret attributes as propositional variables

- an element \( a \in L \) of a lattice \((L, \sqsubseteq, \sqcup, \sqcap, 0)\) with bottom element 0 is called join-irreducible iff \( a \neq 0 \) and if \( a = b \sqcup c \) holds for any \( b, c \in L \), then \( a = b \) or \( a = c \)

- let \( B(N) \) denote the join-irreducibles of \((Sub(N), \leq, \sqcup, \sqcap, \lambda_N)\)

- for \( N = \text{Soccer}\{\text{Match}(\text{Winner},\text{Loser})\} \) we have \( \text{Soccer}\{\text{Match}(\lambda,\lambda)\} \), \( \text{Soccer}\{\text{Match}(\text{Winner},\lambda)\} \) and \( \text{Soccer}\{\text{Match}(\lambda,\text{Loser})\} \) in \( B(N) \)

- can’t express \( \text{Soccer}\{\text{Match}(\text{Winner},\lambda)\} \rightarrow \text{Soccer}\{\text{Match}(\text{Winner},\text{Loser})\} \) since it is different from
\[
\text{Soccer}\{\text{Match}(\text{Winner},\lambda)\} \rightarrow \{\text{Soccer}\{\text{Match}(\text{Winner},\lambda)\}, \text{Soccer}\{\text{Match}(\lambda,\text{Loser})\}\}
\]
4.2 Extended Join-Irreducibles

- *extended join-irreducibles* form smallest $\mathcal{E}(N) \subseteq \text{Sub}(N)$ such that
  (i) $\mathcal{B}(N) \subseteq \mathcal{E}(N)$, and
  (ii) for all $X, Y \in \mathcal{E}(N)$ which are not reconcilable also $X \cup Y \in \mathcal{E}(N)$

- FDs are $\mathcal{X} \rightarrow \mathcal{Y}$ with $\leq$-antichains $\mathcal{X}, \mathcal{Y} \subseteq \mathcal{E}(N)$

- interpret extended join-irreducibles as variables via $\psi : \mathcal{E}(N) \rightarrow \mathcal{V}$

- $\sigma = \{X_1, \ldots, X_n\} \rightarrow \{Y_1, \ldots, Y_m\}$ gives set $\Phi(\sigma)$ of Horn clauses
  \[ \bigwedge_{i=1}^{k} \psi(X_i) \Rightarrow \psi(Y_1), \ldots, \bigwedge_{i=1}^{k} \psi(X_i) \Rightarrow \psi(Y_m) \]

- Horn clauses can also encode the structure of $N$
  \[ \Pi_N = \{ \psi(U) \Rightarrow \psi(V) \mid U, V \in \mathcal{E}(N), U \text{ covers } V \} \]
4.3 The Equivalence

Theorem 2. Let $N$ be a nested attribute, $\Sigma$ a set of FDs and $\sigma$ a single FD on $N$. Let $\Pi_N$ denote the Horn clauses which encode the structure of $N$, and $\Pi$ denote the corresponding set of Horn clauses for $\Sigma$. Then

(i) $\Sigma$ implies $\sigma$,
(ii) $\Sigma$ implies $\sigma$ in the world of two-tuple instances, and
(iii) $\Pi \cup \Pi_N$ logically implies $\pi$ for all $\pi \in \Phi(\sigma)$

are equivalent.

- this extends a well-known result by Fagin et al. (1977), where
  - only single application of record constructor allowed,
  - join-irreducibles form anti-chain, and
  - join-irreducibles (attributes) suffice
4.4 A simple Example

• bijection $\psi$:

\[
\begin{align*}
\text{Cup}(\text{Day}, \lambda) & \leftrightarrow V_0, \\
\text{Cup}(\lambda, \text{Soccer}\{\text{Match}(\lambda, \lambda)\}) & \leftrightarrow V_1, \\
\text{Cup}(\lambda, \text{Soccer}\{\text{Match}(\text{Winner}, \lambda)\}) & \leftrightarrow V_2, \\
\text{Cup}(\lambda, \text{Soccer}\{\text{Match}(\lambda, \text{Loser})\}) & \leftrightarrow V_3, \\
\text{Cup}(\lambda, \text{Soccer}\{\text{Match}(\text{Winner}, \text{Loser})\}) & \leftrightarrow V_4
\end{align*}
\]

• $\text{Cup}(\text{Day}, \lambda) \rightarrow \text{Cup}(\lambda, \text{Soccer}\{\text{Match}(\text{Winner}, \text{Loser})\})$ implies $\text{Cup}(\text{Day}, \lambda) \rightarrow \{\text{Cup}(\lambda, \text{Soccer}\{\text{Match}(\lambda, \text{Winner})\}), \text{Cup}(\lambda, \text{Soccer}\{\text{Match}(\lambda, \text{Loser})\})\}$

• equivalent to $\{V_0 \Rightarrow V_4, V_4 \Rightarrow V_3, V_4 \Rightarrow V_2, V_2 \Rightarrow V_1, V_3 \Rightarrow V_1\}$ implies $V_0 \Rightarrow V_2$ and $V_0 \Rightarrow V_3$
5.1 PO-Spaces

- remove structural rules, implementation details of implication problem

- topological space $\mathcal{T}$ is a structure $(S, \mathcal{C})$ with set $S$ and operation $\mathcal{C}$ mapping subsets of $S$ to subsets of $S$ such that for all $A, B \subseteq S$:
  - $A \subseteq \mathcal{C}A$,
  - $\mathcal{C}A = \mathcal{C}\mathcal{C}A$,
  - $\mathcal{C}(A \cup B) = \mathcal{C}A \cup \mathcal{C}B$, and
  - $\mathcal{C}\emptyset = \emptyset$.

- subset $A$ of $S$ is closed iff $\mathcal{C}A = A$

- poset $(S, \leq)$, for $A \subseteq S$: $\mathcal{C}A = \{b \in S \mid b \leq a \text{ for some } a \in A\}$

- topological space $(S, \mathcal{C})$ is called a PO-space
5.2 Units

- reduce notion of reconcilability to comparability wrt $\leq$
- $U \in \text{Sub}(N)$ is a unit of $N$ iff $U$ is $\leq$-maximal with
  \[ \forall X, Y \leq U \text{ if } X \text{ and } Y \text{ are reconcilable, then } X \leq Y \text{ or } Y \leq X \]
- units of $\text{Cup}(\text{Day}, \text{Soccer}\{\text{Match}(\text{Winner}, \text{Loser})\})$ are $\text{Cup}(\text{Day}, \lambda)$ and $\text{Cup}(\lambda, \text{Soccer}\{\text{Match}(\text{Winner}, \text{Loser})\})$
- $V, W \in \text{Sub}(N)$ reconcilable iff
  for all units $U$ of $N$: $V \sqcap U$ and $W \sqcap U$ are $\leq$-comparable
- $\text{Cup}(\lambda, \text{Soccer}\{\text{Match}(\text{Winner}, \lambda)\})$ and $\text{Cup}(\lambda, \text{Soccer}\{\text{Match}(\lambda, \text{Loser})\})$ are not comparable
5.3 Topological View on FDs

- \( \mathcal{U}(N) = \{U_1, \ldots, U_k\} \), FD on \( N \) is \( \mathcal{X} \rightarrow \mathcal{Y} \) where
  - \( \mathcal{X} = (\mathcal{X}_1, \ldots, \mathcal{X}_k) \), \( \mathcal{Y} = (\mathcal{Y}_1, \ldots, \mathcal{Y}_k) \) and
  - \( \mathcal{X}_i, \mathcal{Y}_i \) are closed sets of the PO-space on \( (\mathcal{E}(U_i), \leq) \)

- \( \models_r \mathcal{X} \rightarrow \mathcal{Y} \) on \( N \) iff \( \forall t_1, t_2 \in r \) we have
  \( \forall i. \forall X \in \mathcal{X}_i. \pi_X^N(t_1) = \pi_X^N(t_2) \) implies \( \forall i. \forall Y \in \mathcal{Y}_i. \pi_Y^N(t_1) = \pi_Y^N(t_2) \)

- \( \mathcal{Y} \subseteq \mathcal{X} \) iff \( \forall i. \mathcal{Y}_i \subseteq \mathcal{X}_i \), and \( \mathcal{X} \cup \mathcal{Y} = (\mathcal{X}_1 \cup \mathcal{Y}_1, \ldots, \mathcal{X}_k \cup \mathcal{Y}_k) \)

**Theorem 3.**

\[
\begin{align*}
\mathcal{X} \rightarrow \mathcal{Y} & \quad \mathcal{Y} \subseteq \mathcal{X} \\
\mathcal{X} \rightarrow \mathcal{Y} & \quad \mathcal{X} \rightarrow \mathcal{X} \cup \mathcal{Y} \\
\mathcal{X} \rightarrow \mathcal{Y}, \mathcal{Y} \rightarrow \mathcal{Z} & \quad \mathcal{X} \rightarrow \mathcal{Z}
\end{align*}
\]

are minimal, sound and complete for FD-implication \( \square \)
6 Conclusion and Future Work

- framework of nested attributes allows to capture data models by including corresponding type constructors
- theory of Brouwerian algebras can be used to extend many achievements from relational databases
- allows to study direct impact of type constructor on design problem without considering peculiarities of specific data model
- study different classes of dependencies in different combinations of constructors
- increase expressiveness by studying embedded dependencies (allowing several Brouwerian algebras simultaneously)
- normal forms