Entailment of Constraints

Dr Paolo Guagliardo

dbs-lecturer@ed.ac.uk

THE UNIVERSITY *of* EDINBURGH
informatics

Fall 2018

Implication of constraints

A set Σ of constraints **implies** (or **entails**) a constraint ϕ if

every instance that satisfies Σ also satisfies ϕ

Notation: $\Sigma \models \phi$

Implication problem

Given Σ and ϕ, does Σ imply ϕ ?

Important because

- We never get the list of all constraints that hold in a database
- The given constraints may look fine, but imply some bad ones
- The given constraints may look bad, but imply only good ones

Axiomatization of constraints

Set of rules (**axioms**) to derive constraints

Sound every derived constraint is implied

Complete every implied constraint can be derived

Sound and complete axiomatization gives a procedure $\vdash$ such that

$$\Sigma \models \phi \quad \textbf{if and only if} \quad \Sigma \vdash \phi$$

Notation

Attributes are denoted by A, B, C, ...

If A and B are attributes, AB denotes the set $\{A, B\}$

Sets of attributes are denoted by X, Y, Z, ...

If X and Y are sets of attributes, XY denotes their union $X \cup Y$

If X is a set of attributes and A is an attribute,

XA denotes $X \cup \{A\}$

Armstrong's axioms

Sound and complete axiomatization for FDs

	Essential axioms
Reflexivity:	If $Y \subseteq X$, then $X \to Y$
Augmentation:	If $X \to Y$, then $XZ \to YZ$ for any Z
Transitivity:	If $X \to Y$ and $Y \to Z$, then $X \to Z$
	Other axioms
Union:	If $X \to Y$ and $X \to Z$, then $X \to YZ$
Decomposition:	If $X \to YZ$, then $X \to Y$ and $X \to Z$

Closure of a set of FDs

Let F be a set of FDs
The closure F^+ of F is the set of all FDs implied by the FDs in F

Can be computed using Armstrong's axioms

Example
Closure of $\{A \to B, B \to C\}$ (blackboard)

Attribute closure

The closure $C_F(X)$ of a set X of attributes w.r.t. a set F of FDs is the set of attributes we can derive from X using the FDs in F (i.e., all the attributes A such that $F \vdash X \to A$)

Properties

- $X \subseteq C_F(X)$
- If $X \subseteq Y$, then $C_F(X) \subseteq C_F(Y)$
- $C_F\big(C_F(X)\big) = C_F(X)$

Solution to the implication problem:

$$F \models Y \to Z \quad \textbf{if and only if} \quad Z \subseteq C_F(Y)$$

Closure algorithm

Input: a set F of FDs, and a set X of attributes

Output: $C_F(X)$, the closure of X w.r.t. F

1. unused := F
2. closure := X
3. **while** ($(Y \to Z) \in$ unused and $Y \subseteq$ closure)

   closure := closure $\cup Z$
   unused := unused $- \{Y \to Z\}$
4. **return** closure

Example

Closure of A w.r.t. $\{AB \to C, A \to B, CD \to A\}$ (blackboard)

Keys, candidate keys, and prime attributes

Let R be a relation with set of attributes U and FDs F

$X \subseteq U$ is a key for R if $F \models X \to U$

Equivalently, X is a key if $C_F(X) = U$ (why?)

Candidate keys

Keys X such that, for each $Y \subset X$, Y is not a key
Intuitively, keys with a minimal set of attributes

Prime attribute: an attribute of a candidate key

Attribute closure and candidate keys

Given a set F of FDs on attributes U,
how do we compute all candidate keys?

1. ck $:= \varnothing$
2. $G :=$ DAG of the powerset 2^U of U
   - Nodes are elements of 2^U (sets of attributes)
   - There is an edge from X to Y if $X - Y = \{A\}$
3. Repeat until G is empty:

   Find a node X without children
   if $C_F(X) = U$:
       ck := ck $\cup \{X\}$
       Delete X and all its ancestors from G
   else:
       Delete X from G

Implication of INDs

Given a set of INDs, what other INDs can we infer from it?

Axiomatization

Reflexivity: $R[X] \subseteq R[X]$

Transitivity: If $R[X] \subseteq S[Y]$ and $S[Y] \subseteq T[Z]$, then $R[X] \subseteq T[Z]$

Projection: If $R[X, Y] \subseteq S[W, Z]$ with $|X| = |W|$,
then $R[X] \subseteq S[W]$

Permutation: If $R[A_1, \ldots, A_n] \subseteq S[B_1, \ldots, B_n]$,
then $R[A_{i_1}, \ldots, A_{i_n}] \subseteq S[B_{i_1}, \ldots, B_{1_n}]$,
where $i_1, \ldots, i_n$ is a permutation of $1, \ldots, n$

Sound and complete derivation procedure for INDs

FDs and INDs together

Given a set F of FDs and an FD f, we can decide whether $F \models f$

Given a set G of INDs and an IND g, we can decide whether $G \models g$

What about $F \cup G \models f$ or $F \cup G \models g$?

This problem is **undecidable**: no algorithm can solve it

What if we consider only keys and foreign keys?

The implication problem is still **undecidable**

Unary inclusion dependencies (UINDs)

INDs of the form $R[A] \subseteq S[B]$ where A, B are attributes

The implication problem for FDs and UINDs is decidable in PTIME

Further reading

Abiteboul, Vianu, Hull. **Foundations of Databases**. Addison-Wesley, 1995

Chapter 8 Functional Dependencies

Chapter 9 Inclusion Dependencies

- Algorithm for checking implication of INDs
- Proof that implication of INDs is PSPACE-complete
- Undecidability proof for implication of FDs+INDs
- Axiomatization for FDs+UINDs