Manipulating Functional Dependencies

Dr. Bing Zhou

Closure of FD sets

- Given a relation schema R and set S of FDs
  - is the FD *F* logically implied by S?
- Example
  - R = {A,B,C,G,H,I}
  - S = A $\rightarrow$B, A $\rightarrow$ C, CG $\rightarrow$ H, CG $\rightarrow$ I, B $\rightarrow$ H
  - would A $\rightarrow$ H be logically implied?
  - yes (you can prove this, using the definition of FD)
- Closure of S: S^+ = all FDs logically implied by S
- How to compute S^+ ?
  - we can use <u>Armstrong's axioms</u>

Armstrong's Axioms

- Reflexivity rule
  - $A_1\ A_2 \ldots A_n \rightarrow$ *a subset of* $A_1\ A_2 \ldots A_n$
- Augmentation rule
  - $A_1\ A_2 \ldots A_n \rightarrow B_1\ B_2 \ldots B_m$

    then

    $A_1\ A_2 \ldots A_n\ \ C_1\ C_2 \ldots C_k \rightarrow B_1\ B_2 \ldots B_m\ C_1\ C_2 \ldots C_k$
- Transitivity rule
  - $A_1\ A_2 \ldots A_n \rightarrow B_1\ B_2 \ldots B_m$ and

    $B_1\ B_2 \ldots B_m \rightarrow C_1\ C_2 \ldots C_k$

    then

    $A_1\ A_2 \ldots A_n \rightarrow C_1\ C_2 \ldots C_k$

Inferring S^+ using Armstrong's Axioms

- S^+ = S
- Loop
  - For each *F* in S, apply reflexivity and augmentation rules
  - add the new FDs to S^+
  - For each pair of FDs in S, apply the transitivity rule
  - add the new FD to S^+
- Until S^+ does not change any further

Additional Rules

- Union rule
  - X → Y and X → Z, then X → YZ
  - (X, Y, Z are sets of attributes)

- Decomposition rule
  - X → YZ, then X → Y and X → Z

- Pseudo-transitivity rule
  - X → Y and YZ → U, then XZ → U

- These rules can be inferred from Armstrong's axioms

Example

- $R = (A, B, C, G, H, I)$
  $F = \{ A \to B \quad A \to C \quad CG \to H \quad CG \to I \quad B \to H\}$
- some members of F^+
  - $A \to H$
    - by transitivity from $A \to B$ *and* $B \to H$
  - $AG \to I$
    - by augmenting $A \to C$ with G, to get $AG \to CG$
      and then transitivity with $CG \to I$
  - $CG \to HI$
    - from $CG \to H$ *and* $CG \to I$: "union rule" can be inferred from
      - definition of functional dependencies, or
      - Augmentation of $CG \to I$ to infer $CG \to CGI$, augmentation of $CG \to H$ to infer $CGI \to HI$, and then transitivity

Closures of Attributes

Suppose a relation with attributes A, B, C, D, E, and F satisfies the FDs

$$AB \rightarrow C \quad BC \rightarrow AD \quad D \rightarrow E, \quad CF \rightarrow B$$

Given these FDs,

- what is the set X of attributes such that $AB \rightarrow X$ is true?
  $X = \{A, B, C, D, E\}$, i.e., $AB \rightarrow ABCDE$.

- what is the set Y of attributes such that $BCF \rightarrow Y$ is true?
  $Y = \{A, B, C, D, E, F\}$, i.e., $BCF \rightarrow ABCDEF$

- $\{B, C, F\}$ is a superkey.

Closures of Attributes: Definition

Given

- a set of attributes $\{A_1, A_2, \ldots, A_n\}$ and
- a set of FDs S,

the *closure* of $\{A_1, A_2, \ldots, A_n\}$ under the FDs in S is

- the set of attributes $\{B_1, B_2, \ldots, B_m\}$ such that for $1 \leq i \leq m$, the FD $A_1A_2 \ldots A_n \rightarrow B_i$ follows from S.
- the closure is denoted by $\{A_1, A_2, \ldots, A_n\}^+$.

- Which attributes must $\{A_1, A_2, \ldots, A_n\}^+$ contain at a minimum? $\{A_1, A_2, \ldots, A_n\}$. Why?

$A_1A_2 \ldots A_n \rightarrow A_i$ is a trivial FD.

Closures of Attributes: Algorithm

Given

- a set of attributes $\{A_1, A_2, \ldots, A_n\}$ and
- a set of FDs S,
- compute $X = \{A_1, A_2, \ldots, A_n\}^+$.

1. Set $X \leftarrow \{A_1, A_2, \ldots, A_n\}$.
2. Find an FD $B_1 B_2 \ldots B_k \rightarrow C$ in S such that $\{B_1, B_2, \ldots B_k\} \subseteq X$ but $C \notin X$.
3. Add C to X.
4. Repeat the last two steps until you cannot find such an attribute C.
5. The final value of X is the desired closure.

Closures of Attributes: Algorithm

- Basis: $Y^+ = Y$
- Induction: Look for an FD's left side X that is a subset of the current Y^+
  - If the FD is $X \rightarrow A$, add A to Y^+

Closure Example

Student(SSN, sName, address,
HScode, HSname, HScity, GPA, priority)

SSN → sName, address, GPA
GPA → priority
HScode → HSname, HScity

$\{SSN, HScode\}^+$

{SSN, HScode, sName, address,
GPA, priority, HSname, HScity}

Why is the Concept of Closures Useful?

- Closures allow us to prove correctness of rules for manipulating FDs.
  - *Transitive rule*: if
    $A_1A_2 \ldots A_n \rightarrow B_1B_2 \ldots B_m$
    and
    $B_1B_2 \ldots B_m \rightarrow C_1C_2 \ldots C_n$
    then
    $A_1A_2 \ldots A_n \rightarrow C_1C_2 \ldots C_n$.
  - To prove this rule, simply check if
    $\{C_1, C_2, \ldots, C_n\} \subseteq \{A_1, A_2, \ldots, A_n\}^+$.
- Closures allow us to procedurally define keys. A set of attributes X is a key for a relation R if and only if
  - $\{X\}^+$ is the set of all attributes of R and
  - for no attribute $A \in X$ is $\{X - \{A\}\}^+$ the set of all attributes of R.

Uses of Attribute Closure

There are several uses of the attribute closure algorithm:

- Testing for superkey:
  - To test if α is a superkey, we compute α^{+}, and check if α^{+} contains all attributes of *R*.
- Testing functional dependencies
  - To check if a functional dependency $\alpha \rightarrow \beta$ holds (or, in other words, is in F^{+}), just check if $\beta \subseteq \alpha^{+}$.
  - That is, we compute α^{+} by using attribute closure, and then check if it contains β.
  - Is a simple and cheap test, and very useful
- Computing closure of F
  - For each $\gamma \subseteq R$, we find the closure γ^{+}, and for each $S \subseteq \gamma^{+}$, we output a functional dependency $\gamma \rightarrow S$.

- Does F = {A $\rightarrow$ B, B $\rightarrow$ C, C D $\rightarrow$ E } imply A $\rightarrow$ E?
  - i.e, is A $\rightarrow$ E in the closure F^+ ? Equivalently, is E in A^+?

Example of Attribute Set Closure

- $R = (A, B, C, G, H, I)$
- $F = \{A \rightarrow B \quad A \rightarrow C \quad CG \rightarrow H \quad CG \rightarrow I \quad B \rightarrow H\}$
- $(AG)^+$
  1. $result = AG$
  2. $(A \rightarrow C$ and $A \rightarrow B)$ $result = ABCG$
  3. $(CG \rightarrow H$ and $CG \subseteq AGBC)$ $result = ABCGH$
  4. $(CG \rightarrow I$ and $CG \subseteq AGBCH)$ $result = ABCGHI$

- Is AG a super key?
- Is AG a key?
  1. Does $A^+ \rightarrow R$?
  2. Does $G^+ \rightarrow R$?

Example of Closure Computation

- Consider the "bad" relation `Students(Id, Name, AdvisorId, AdvisorName, FavouriteAdvisorId)`.
- What are the FDs that hold in this relation?

  $\texttt{Id} \rightarrow \texttt{Name}$

  $\texttt{Id} \rightarrow \texttt{FavouriteAdvisorId}$

  $\texttt{AdvisorId} \rightarrow \texttt{AdvisorName}$

- To compute the key for this relation,
  1. Compute the closures for all sets of attributes.
  2. Find the minimal set of attributes whose closure is the set of all attributes.

Closures of FDs vs. Closures of Attributes

- Both algorithms take as input a relation R and a set of FDs F.
- Closure of FDs:
  - Computes $\{F\}^+$, the set of all FDs that follow from F.
  - Output is a set of FDs.
  - Output may contain an exponential number of FDs.
- Closure of attributes:
  - In addition, takes a set $\{A_1, A_2, \ldots, A_n\}$ of attributes as input.
  - Computes $\{A_1, A_2, \ldots, A_n\}^+$, the set of all attributes B such that the $A_1A_2 \ldots A_n \rightarrow B$ follows from F.
  - Output is a set of attributes.
  - Output may contain at most the number of attributes in R.