Relational Design Theory

Carla Teixeira Lopes

Bases de Dados
Mestrado Integrado em Engenharia Informática e Computação, FEUP

Based on Jennifer Widom and Christopher Ré slides

Agenda

~~Relational Design Overview~~

~~Functional Dependencies~~

Closures, Superkeys and Keys

Inferring Functional Dependencies

Normal Forms

Decompositions

Closure of attributes

Given a relation, FDs, set of attributes $\bar{A}$, find all B such that $\bar{A} \rightarrow B$

$\bar{A}^{+}$ is the closure of $\bar{A}$

- Finding the set of attributes functionally determined by $\{A_1, \ldots, A_n\}^{+}$

Algorithm

- Start with $\{A_1, \ldots, A_n\}$
- Repeat until no change:
  - If $\bar{X} \rightarrow \bar{Y}$ and $\bar{X}$ in set, add $\bar{Y}$ to set

Applying combining and transitive rules

Closure example

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

FD1. SSN -> sName, address, GPA

FD2. GPA -> priority

FD3. HScode -> HSname, HScity

Compute $\{SSN, HScode\}^+$

{SSN, HScode}

FD1

{SSN, Hscode, sName, address, GPA}

FD2

{SSN, Hscode, sName, address, GPA, priority}

FD3

{SSN, Hscode, sName, address, GPA, priority, Hsname, HScity}

Closure example

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

FD1. SSN -> sName, address, GPA

FD2. GPA -> priority

FD3. HScode -> HSname, HScity

$\{SSN, HScode\}^{+}$

{SSN, Hscode, sName, address, GPA, priority, Hsname, HScity}

All attributes of Student

Key for the relation

Exercise

R (A, B, C, D, E, F)

{A,B} -> {C}

{A,D} -> {E}

{B} -> {D}

{A,F} -> {B}

Compute $\{A, B\}^+$

Compute $\{A, F\}^+$

Exercise

R (A, B, C, D, E, F)

{A,B} -> {C}

{A,D} -> {E}

{B} -> {D}

{A,F} -> {B}

Compute $\{A, B\}^{+}$= {A, B}

Compute $\{A, F\}^{+}$ = {A, F}

Exercise

R (A, B, C, D, E, F)

{A,B} -> {C}

{A,D} -> {E}

{B} -> {D}

{A,F} -> {B}

Compute $\{A, B\}^+$= {A, B, C, D}

Compute $\{A, F\}^+$ = {A, F, B}

Exercise

R (A, B, C, D, E, F)

{A,B} -> {C}

{A,D} -> {E}

{B} -> {D}

{A,F} -> {B}

Compute $\{A, B\}^+$= {A, B, C, D, E}

Compute $\{A, F\}^+$ = {A, F, B, C, D}

Exercise

R (A, B, C, D, E, F)

{A,B} -> {C}

{A,D} -> {E}

{B} -> {D}

{A,F} -> {B}

Compute $\{A, B\}^{+}$= {A, B, C, D, E}

Compute $\{A, F\}^{+}$ = {A, F, B, C, D, E}

Closure and keys

Is $\bar{A}$ a key for a relation R with a set of FDs?

If $\bar{A}^{+}$ contains all attributes of R, then $\bar{A}$ is a key

How can we find all keys given a set of FDs?

Consider every subset of attributes and compute its closure to see if it determines all attributes

To increase efficiency, consider the subsets in increasing order

If AB is a key, $AB -\!\!>$ all attributes, every superset of AB is also a key

Start with single attributes, then go to pairs and so on

Find all keys - Example

Compute X^+, for every set of attributes X in R (A,B,C,D)

{A,B} -> {C}

{A, D} -> {B}

{B} -> {D}

$\{A\}^+ = \{A\}$
$\{B\}^+ = \{B,D\}$
$\{C\}^+ = \{C\}$
$\{D\}^+ = \{D\}$

Find all keys - Example

Compute X^+, for every set of attributes X in R (A,B,C,D)

{A,B} -> {C}
{A, D} -> {B}
{B} -> {D}

$\{A\}^+ = \{A\}$
$\{B\}^+ = \{B,D\}$
$\{C\}^+ = \{C\}$
$\{D\}^+ = \{D\}$
$\{A,B\}^+ = \{A,B,C,D\}$
$\{A,C\}^+ = \{A,C\}$
$\{A,D\}^+ = \{A,B,C,D\}$
$\{B,C\}^+ = \{B,C,D\}$
$\{B,D\}^+ = \{B,D\}$
$\{C,D\}^+ = \{C,D\}$

Find all keys - Example

Compute X^+, for every set of attributes X in R (A,B,C,D)

$\{A,B\} \to \{C\}$

$\{A, D\} \to \{B\}$

$\{B\} \to \{D\}$

$\{A\}^+ = \{A\}$
$\{B\}^+ = \{B,D\}$
$\{C\}^+ = \{C\}$
$\{D\}^+ = \{D\}$
$\{A,B\}^+ = \{A,B,C,D\}$
$\{A,C\}^+ = \{A,C\}$
$\{A,D\}^+ = \{A,B,C,D\}$
$\{B,C\}^+ = \{B,C,D\}$
$\{B,D\}^+ = \{B,D\}$
$\{C,D\}^+ = \{C,D\}$
$\{A,B,C\}^+ = \{A,B,D\}^+ = \{A,C,D\}^+ = \{A,B,C,D\}$
$\{B,C,D\}^+ = \{B,C,D\}$

Don’t need to compute

Find all keys - Example

Compute X^+, for every set of attributes X in R (A,B,C,D)

{A,B} -> {C}

{A, D} -> {B}

{B} -> {D}

$\{A\}^+ = \{A\}$
$\{B\}^+ = \{B,D\}$
$\{C\}^+ = \{C\}$
$\{D\}^+ = \{D\}$
$\{A,B\}^+ = \{A,B,C,D\}$
$\{A,C\}^+ = \{A,C\}$
$\{A,D\}^+ = \{A,B,C,D\}$
$\{B,C\}^+ = \{B,C,D\}$
$\{B,D\}^+ = \{B,D\}$
$\{C,D\}^+ = \{C,D\}$
$\{A,B,C\}^+ = \{A,B,D\}^+ = \{A,C,D\}^+ = \{A,B,C,D\}$
$\{B,C,D\}^+ = \{B,C,D\}$
$\{A,B,C,D\}^+ = \{A,B,C,D\}$ → Don't need to compute

Find all keys - Example

Compute X^+, for every set of attributes X in R (A,B,C,D)

{A,B} -> {C}
{A, D} -> {B}
{B} -> {D}

$\{A\}^+ = \{A\}$
$\{B\}^+ = \{B,D\}$
$\{C\}^+ = \{C\}$
$\{D\}^+ = \{D\}$
$\{A,B\}^+ = \{A,B,C,D\}$
$\{A,C\}^+ = \{A,C\}$
$\{A,D\}^+ = \{A,B,C,D\}$
$\{B,C\}^+ = \{B,C,D\}$
$\{B,D\}^+ = \{B,D\}$
$\{C,D\}^+ = \{C,D\}$
$\{A,B,C\}^+ = \{A,B,D\}^+ = \{A,C,D\}^+ = \{A,B,C,D\}$
$\{B,C,D\}^+ = \{B,C,D\}$
$\{A,B,C,D\}^+ = \{A,B,C,D\}$

(Super)keys

{A,B}
{A,D}
{A,B,C}
{A,B,D}
{A,C,D}
{A,B,C,D}

Superkeys and keys

A superkey is a set of attributes $A_1, \ldots, A_n$ such that for *any other* attribute B in R, we have $\{A_1, \ldots, A_n\} \rightarrow B$

all attributes are functionally determined by a superkey

A key is a minimal superkey

Meaning that no subset of a key is also a superkey

Also named *candidate key*

Primary key is one and **only one** of the keys

Example of finding keys

Product (name, price, category, color)

{name, category} → price

{category} → color

What is the key?

Example of finding keys

Product (name, price, category, color)

{name, category} → price

{category} → color

{name, category}$^{+}$ = {name, price, category, color}

= the set of all attributes

this is a **superkey**

this is a **key**, since neither *name* nor *category* alone is a superkey

Agenda

~~Relational Design Overview~~

~~Functional Dependencies~~

~~Closures, Superkeys and Keys~~

Inferring Functional Dependencies

Normal Forms

Decompositions

Inferring FDs

S1 and S2 sets of FDs

S2 “follows from” S1 if every relation instance satisfying S1 also satisfies S2

Example

S2: {SSN->priority}

S1: {SSN->GPA, GPA->priority}

Inferring FDs

How to test if $\bar{A}$ -> $\bar{B}$ follows from S ?

Compute $\bar{A}^{+}$ based on S and check if $\bar{B}$ is in set

Armstrong's Axioms

Set of rules that are what's called complete

If one thing about functional dependencies can be proved from another, then it can be proved using the Armstrong's Axioms

Goal: Find minimal set of completely nontrivial FDs such that all FDs that hold on the relation follow from the dependencies in this set

Using Closure to Infer ALL FDs

Given F = $\boxed{\{A,B\} \rightarrow C,\ \{A,D\} \rightarrow B,\ \{B\} \rightarrow D}$

Step 1: Compute X^+, for every set of attributes X:

$\{A\}^+ = \{A\}$, $\{B\}^+ = \{B,D\}$, $\{C\}^+ = \{C\}$, $\{D\}^+ = \{D\}$, $\{A,B\}^+ = \{A,B,C,D\}$, $\{A,C\}^+ = \{A,C\}$, $\{A,D\}^+ = \{A,B,C,D\}$, $\{A,B,C\}^+ = \{A,B,D\}^+ = \{A,C,D\}^+ = \{A,B,C,D\}$, $\{B,C,D\}^+ = \{B,C,D\}$, $\{A,B,C,D\}^+ = \{A,B,C,D\}$

Y in the closure of X

Step 2: Enumerate all FDs $X \rightarrow Y$, such that $Y \subseteq X^+$ and $X \cap Y = \varnothing$:

X →Y is non-trivial

$\{B\} \rightarrow \{D\}$, $\{A,B\} \rightarrow \{C,D\}$, $\{A,D\} \rightarrow \{B,C\}$, $\{A,B,C\} \rightarrow \{D\}$, $\{A,B,D\} \rightarrow \{C\}$, $\{A,C,D\} \rightarrow \{B\}$

Projecting a set of FDs

Input: relation R; FDs for R; relation $R_1 = \pi_L(R)$

Output: T, the set of FDs that hold in R_1

For each set of attributes $\bar{X}$ of R_1, compute $\bar{X}^+$

With respect to the FDs for R. These FDs may involve attributes that are in R and not in R_1

$\bar{X} \cap \bar{A} = \varnothing$

Add to T all nontrivial FDs $\bar{X}$ -> $\bar{A}$, such that $\bar{X}^+ \supseteq \bar{A}$ and $\bar{A}$ contains attributes of R_1

For minimal base, repeat until no changes

Remove FDs from T that follow from the other FDs in T

Replace $\bar{Y} \rightarrow \bar{B}$ by $\bar{Z} \rightarrow \bar{B}$ if $\bar{Z}$ is $\bar{Y}$ with one of its attributes removed

Projecting a set of FDs Example

Input: R (A, B, C, D); FDs: A->B, B->C, C->D; R_1= (A, C, D)

Compute the closure for all the subsets of {A, C, D}

- $\{A\}^+$={A, B, C, D} thus A->C and A->D hold in R1
  - No need to consider any superset of {A}, every FD would follow an FD with only A on the left side (e.g.: AC->D follows from A->D)
- $\{C\}^+$={C, D} thus C->D holds in R1
- $\{D\}^+$={D}
- $\{C,D\}^+$={C,D}

If $\bar{X}$ is a key of R1, no need to close supersets of $\bar{X}$

No need to close the empty set and the set of all attributes

- Cannot yield a nontrivial FD

Minimal base: A->C and C->D

Kahoot time!

Any doubts?

Readings

Jeffrey Ullman, Jennifer Widom, A first course in Database Systems 3rd Edition

- Section 3.1 – Functional Dependencies
- Section 3.2 – Rules About Functional Dependencies
- Section 3.3 – Design of Relational Database Schemas
- Section 3.4 – Decomposition: The Good, Bad, and Ugly
- Section 3.5 – Third Normal Form