Database Management Systems (DBMS)

Lec 17- FDs: Inference Rules, Equivalence, and Minimal Cover

Ramesh K. Jallu
IIIT Raichur

Date: 30/03/21

Recap

- Multivalued dependency: The fourth normal form (4NF)
- The join dependency: The fifth normal form (5NF)

Today's plan

- Functional dependency
    - Inference rules
    - Equivalance
    - Minimal cover

Functional dependencies

- So far we illustrated FDs with some examples, and multiple FDs over a single relation
- We identified and discussed problematic functional dependencies
- They can be eliminated by a proper decomposition of a relation. This process was described as ***normalization***
- We now study of functional dependencies and show how new dependencies can be inferred from a given set of FDs

Inference Rules for FDs

- *Let* $\boldsymbol{F}$ denote the set of functional dependencies that are specified on relation schema $\boldsymbol{R}$

- An FD $\boldsymbol{X} \rightarrow \boldsymbol{Y}$ is ***inferred from*** a set of dependencies $\boldsymbol{F}$ specified on $\boldsymbol{R}$ if $\boldsymbol{X} \rightarrow \boldsymbol{Y}$ holds in *every* legal relation state $\boldsymbol{r}$ of $\boldsymbol{R}$
  - I.e., whenever $\boldsymbol{r}$ satisfies all the dependencies in $\boldsymbol{F}$, $\boldsymbol{X} \rightarrow \boldsymbol{Y}$ also holds in $\boldsymbol{r}$

- The set of all dependencies that include $\boldsymbol{F}$ as well as all dependencies that can be inferred from $\boldsymbol{F}$ is called the ***closure*** of $\boldsymbol{F}$; it is denoted by $\boldsymbol{F+}$

Examples

- $F = \{\text{Dept\_no} \rightarrow \text{Mgr\_ssn}, \text{Mgr\_ssn} \rightarrow \text{Mgr\_phone}\}$
  - $\text{Dept\_no} \rightarrow \text{Mgr\_phone}$
- $F = \{\text{Ssn} \rightarrow \{\text{Ename, Bdate, Address, Dnumber}\},$
  $\text{Dnumber} \rightarrow \{\text{Dname, Dmgr\_ssn}\}\}$
  - $\text{Ssn} \rightarrow \{\text{Dname, Dmgr\_ssn}\}$
  - $\text{Ssn} \rightarrow \text{Ssn}$
  - $\text{Dnumber} \rightarrow \text{Dname}$

Inference Rules for FDs (Contd.)

- The rules we use to infer new dependencies from a given set of dependencies are called ***inference rules***

- We use the notation $F \models X \to Y$ to denote that the functional dependency $X \to Y$ is inferred from the set of functional dependencies F

- The FD $\{X, Y\} \to Z$ is abbreviated to $XY \to Z$
    - The FD $\{X, Y, Z\} \to \{U, V\}$ is abbreviated to $XYZ \to UV$

Armstrong's axioms

- ***Reflexive rule*** (**IR1**): If $X \supseteq Y$, then $X \rightarrow Y$
- ***Augmentation rule*** (**IR2**): $\{X \rightarrow Y\} \models XZ \rightarrow YZ$
- ***Transitive rule*** (**IR3**): $\{X \rightarrow Y, Y \rightarrow Z\} \models X \rightarrow Z$
- Armstrong axioms refer to the *Sound* and *Complete*

Proof of Armstrong's axioms

Proof of IR1. Suppose that $X \supseteq Y$ and that two tuples t_1 and t_2 exist in some relation instance r of R such that $t_1\,[X] = t_2\,[X]$. Then $t_1[Y] = t_2[Y]$ because $X \supseteq Y$; hence, $X \rightarrow$ Y must hold in r.

Proof of IR2 (by contradiction). Assume that $X \rightarrow Y$ holds in a relation instance r of R but that $XZ \rightarrow YZ$ does not hold. Then there must exist two tuples t_1 and t_2 in r such that (1) $t_1\,[X] = t_2\,[X]$, (2) $t_1\,[Y] = t_2\,[Y]$, (3) $t_1\,[XZ] = t_2\,[XZ]$, and (4) $t_1\,[YZ] \neq t_2\,[YZ]$. This is not possible because from (1) and (3) we deduce (5) $t_1\,[Z] = t_2\,[Z]$, and from (2) and (5) we deduce (6) $t_1\,[YZ] = t_2\,[YZ]$, contradicting (4).

Proof of IR3. Assume that (1) $X \rightarrow Y$ and (2) $Y \rightarrow Z$ both hold in a relation r. Then for any two tuples t_1 and t_2 in r such that $t_1\,[X] = t_2\,[X]$, we must have (3) $t_1\,[Y] = t_2\,[Y]$, from assumption (1); hence we must also have (4) $t_1\,[Z] = t_2\,[Z]$ from (3) and assumption (2); thus $X \rightarrow Z$ must hold in r.

Secondary axioms

- ***Decomposition rule*** (**IR4**): $\{X \rightarrow YZ\} \models X \rightarrow Y, X \rightarrow Z$
  - $X \rightarrow \{A_1, A_2, ..., A_n\} \models \{X \rightarrow A_1, X \rightarrow A_2, ..., X \rightarrow A_n\}$
- ***Additive*** (or) ***Union rule*** (**IR5**): $\{X \rightarrow Y, X \rightarrow Z\} \models X \rightarrow YZ$
  - $\{X \rightarrow A_1, X \rightarrow A_2, ... , X \rightarrow A_n\} \models X \rightarrow \{A_1, A_2, ... , A_n\}$
- ***Pseudo-transitive rule*** (**IR6**): $\{X \rightarrow Y, \mathrm{W}Y \rightarrow Z\} \models \mathrm{W}X \rightarrow Z$

Proof of IR5

Proof of IR5 (using IR1 through IR3).

1. $X \rightarrow Y$ (given).
2. $X \rightarrow Z$ (given).
3. $X \rightarrow XY$ (using IR2 on 1 by augmenting with X; notice that $XX = X$).
4. $XY \rightarrow YZ$ (using IR2 on 2 by augmenting with Y).
5. $X \rightarrow YZ$ (using IR3 on 3 and 4).

- *True* or *false*: Justify your answer
  i. $\{X \rightarrow A, Y \rightarrow B\} \models XY \rightarrow AB$
  ii. $XY \rightarrow A \models X \rightarrow A$ or $Y \rightarrow A$

Clouser of a set of attributes

- WKT, from ***F*** we can infer FDs by applying the rules

- A systematic way to determine additional FDs is to determine

  i. each set of attributes ***X*** that appears as a left-hand side of some functional dependency in ***F***

  ii. the set of *all attributes* that are dependent on ***X***

- For each set of attributes ***X***, we determine the set X^+ of attributes that are functionally determined by ***X*** based on ***F***; where ***X+*** is called the ***closure of X under F***

Algorithm to determine X^+

- **Algorithm**:

  - *Input*: A set F of FDs on a relation schema R, and a set of attributes X, which is a subset of R
  - *Output*: X^+

  1. $X^+ := X$;
  2. for each functional dependency $Y \rightarrow Z$ in F do

     if $X^+ \supseteq Y$ then $X^+ := X^+ \cup Z$;

Example

- Consider the following relation schema about classes held at a university in a given academic year

- **CLASS**(Classid, Course_No, Instr_name, Credit_hrs, Text, Publisher, Classroom, Capacity)

- ***F*** = {FD1,FD2,FD3,FD4,FD5}, where
  - FD1: Classid → {Course_No, Instr_name, Credit_hrs, Text, Publisher, Classroom, Capacity}
  - FD2: Course_No → Credit_hrs
  - FD3: {Course_No, Instr_name} → {Text, Classroom}
  - FD4: Text → Publisher
  - FD5: Classroom → Capacity

Example (Contd.)

1. $\{Classid\}^+ = \{Classid, Course\_No, Instr\_name, Credit\_hrs, Text, Publisher, Classroom, Capacity\} = CLASS$

2. $\{Course\_No\}^+ = \{Course\_No, Credit\_hrs\}$

3. $\{Course\_No, Instr\_name\}^+ = \{Course\_No, Instr\_Name, Credit\_hrs, Text, Publisher, Classroom, Capacity\}$

Equivalence between two sets of FDs

- A set of FDs ***F*** is said to ***cover*** another set of FDs ***E*** if every FD in ***E*** is also in F^+
  - I.e., every dependency in ***E*** can be inferred from ***F***
  - Alternatively, we can say that ***E*** is ***covered by F***

- Two sets of FDs ***E*** and ***F*** are equivalent if $E^+ = F^+$
  - I.e., every FD in ***E*** can be inferred from ***F***, and every FD in ***F*** can be inferred from ***E***
  - We say ***E*** is ***equivalent*** to ***F*** if both the conditions ***E*** covers ***F*** and ***F*** covers ***E*** hold

Testing of equivalance

1. We can determine whether F covers E or not
   a. by calculating X^+ wrt F for each FD $X \rightarrow Y$ in E, and then checking whether this X^+ includes the attributes in Y
   b. If this is the case for every FD in E, then F covers E

2. Similarly we can check whether E covers F or not

3. If F covers E and E covers F, then F and E are equivalent

Example

Let $F = \{A \rightarrow C, AC \rightarrow D, E \rightarrow AD, E \rightarrow H\}$ and $G = \{A \rightarrow CD, E \rightarrow AH\}$.
Test whether F and G are equivalent or not

1. We need to check first whether F covers G or not
   i. Consider the FD $A \rightarrow CD$
      - $A^+ = \{A, C, D\}$; A^+ contains the attributes C and D

   ii. Consider the FD $E \rightarrow AH$
      - $E^+ = \{E, H, A, D, C, D\}$; E^+ contains A and H

   iii. We can conclude that F covers G

Example (Contd.)

2. We need to check now whether G covers F or not

   i. Consider the FD $A \rightarrow C$

      - $A^+ = \{A, C, D\}$; A^+ includes the attribute C

   ii. Consider the FD $AC \rightarrow D$

      - $\{A,C\}^+ = \{A, C, D\}$; $\{A,C\}^+$ contains D

   iii. Consider the FD $E \rightarrow AD$

      - $E^+ = \{E, A, H, C, D\}$; E^+ contains A and D

   iv. Consider the FD $E \rightarrow H$

   v. We can conclude that G covers F

$F = \{A \rightarrow C, AC \rightarrow D, E \rightarrow AD, E \rightarrow H\}$

$G = \{A \rightarrow CD, E \rightarrow AH\}$

Minimal Sets of Functional Dependencies

- We apply inference rules on ***F*** to compute its closure F^+
  - I.e., we expand ***F*** to F^+
  - What about the opposite?
    - I.e., can we shrink ***F*** to its minimal form so that the minimal set is still equivalent to the original set ***F***

- A ***minimal cover*** of a set of FDs ***E*** is a set of FDs ***F*** that satisfies the property that every dependency in ***E*** is in the closure F^+ of ***F***
  - In addition, this property is lost if any dependency from the set ***F*** is removed

Example

- Let $E = \{B \rightarrow A,\ D \rightarrow A,\ AB \rightarrow D\}$. The minimal cover of E is $F = \{B \rightarrow D,\ D \rightarrow A\}$

- The closure of F, $F^+ = \{B \rightarrow D,\ D \rightarrow A,\ B \rightarrow A,\ AB \rightarrow D,\ \ldots\}$

Thank you!