Inferring FDs

25

- Given a set of FDs, we can often infer further FDs.
- This will come in handy when we apply FDs to the problem of database design.

25

Dependency Inference

26

- Suppose we are given FDs
  $X_1 \rightarrow A_1$,
  $X_2 \rightarrow A_2$,
  …,
  $X_n \rightarrow A_n$.
- Does the FD $Y \rightarrow B$ also hold in any relation that satisfies the given FDs?
- Example: If $A \rightarrow B$ and $B \rightarrow C$ hold, surely $A \rightarrow C$ holds, even if we don't say so.
  $A \rightarrow C$ is ***entailed (implied)*** by $\{A \rightarrow B, B \rightarrow C\}$

26

Transitive Property

27

The transitive property holds for FDs

- Consider the FDs: $A \rightarrow B$ and $B \rightarrow C$; then $A \rightarrow C$ holds
- Consider the FDs: $AD \rightarrow B$ and $B \rightarrow CD$; then $AD \rightarrow CD$ holds or just $AD \rightarrow C$ (because of trivial FDs)

27

Method 1: Prove it from first principles

28

- To test if $Y \rightarrow B$, start by assuming two tuples agree on all attributes of Y.

← Y →

t1: aaaaa bb. . . b

t2: aaaaa ?? . . . ?

28

Example

29

ClientID	Income	OtherProd	Rate	Country	City	State
225	High	A	2.1%	USA	San Francisco	MD
420	High	A	2.1%	USA	San Francisco	CA
333	High	B	3.0%	USA	San Francisco	CA
576	High	B	3.0%	USA	San Francisco	CA
128	Low	C	4.5%	UK	Reading	Berkshire
193	Low	C	4.5%	UK	London	London
550	Low	B	3.5%	UK	London	London

F1: [Income, OtherProd] → [Rate]   F2: [Country, City] → [State]

How to prove it in the general case?

29

Closure Test for FDs

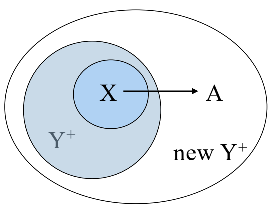
30

- Given attribute set Y and FD set F
  - Denote Y_F^+ or Y^+ the closure of Y relative to F
    Y_F^+ = set of all FDs given or implied by Y
- Computing the closure of Y
  - Start: $Y_F^+ = Y$, $F' = F$
  - While there exists an $f \in F'$ s.t. $LHS(f) \subseteq Y_F^+$:
    $Y_F^+ = Y_F^+ \cup RHS(f)$
    $F' = F' - f$
  - At end: $Y \to B$ for all $B \in Y_F^+$

Acknowledgements: M. Papagelis

30

Computing the closure Y^+ of a set of attributes Y
Given FDs F:

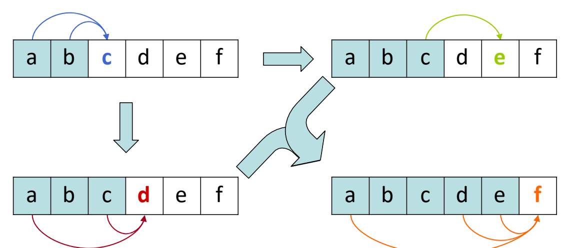


31

Example: Closure Test

32

- Consider R(a,b,c,d,e,f)
  with FDs ab → c, ac → d, c → e, ade → f
- Find Y^+ if Y = ab or find $\{a,b\}^+$



$\{a,b\}^+=\{a,b,c,d,e,f\}$ or *ab → cdef*

ab is a candidate key!

32

Lecture Example

33

33

Lecture Example

34

34

Discarding redundant FDs

35

- Minimal basis: opposite from closure
- Given a set of FDs F, want a minimal F' s.t.
  - $F' \subseteq F$
  - F' entails f for all $f \in F$
- Properties of a minimal basis F'
  - RHS is always singleton
  - If any FD is removed from F', F' is no longer a minimal basis
  - If for any FD in F' we remove one or more attributes from the LHS of $f \in F'$, the result is no longer a minimal basis

35

In other words …

36

- *Minimal basis also referred to as minimal cover*
- *Minimal basis* for FDs:
  - Right sides are single attributes.
  - No FD can be removed.
  - No attribute can be removed from a left side.
- Constructing a minimal cover
  - Decompose RHS to single attributes
  - Repeatedly try to remove an FD and see if remaining FDs are equivalent to original set. That is, does the closure of the LHS attributes (of the removed FD) include the RHS attribute?
  - Repeatedly try to remove an attribute from a LHS and see if the removed attribute can be derived from the remaining FDs.

36

Constructing a minimal basis

37

Straightforward but time-consuming

1. Split all RHS into singletons
2. For all f in F', test whether J = $(F'-f)^+$ is still equivalent to F^+

=> Might make F' too small

3. For all i $\in$ LHS(f), for all f $\in$ F', let LHS(f')=LHS(f)-i
   Test whether $(F'-f+f')^+$ is still equivalent to F^+
4. Repeat (2) and (3) until neither makes progress

37

Minimal Basis: Example

38

- Relation R: R(A, B, C, D)
- Defined FDs:
  - F = {A->AC, B->ABC, D->ABC}

Find the minimal basis M of F

38

Minimal Basis: Example (cont.)

39

F = {A->AC, B->ABC, D->ABC}

1st Step

- H = {~~A->A~~, A->C, B->A, ~~B->B~~, ~~B->C~~, ~~D->A~~, D->B, ~~D->C~~}

2nd Step

- A->A, B->B: can be removed as trivial
- A->C: can't be removed, as there is no other LHS with A
- B->A: can't be removed, because for J=H-{B->A} is B^+=BC
- B->C: can be removed, because for J=H-{B->C} is B^+=ABC
- D->A: can be removed, because for J=H-{D->A} is D^+=DBA
- D->B: can't be removed, because for J=H-{D->B} is D^+=DC
- D->C: can be removed, because for J=H-{D->C} is D^+=DBAC

Step outcome => H = {A->C, B->A, D->B}

39

Minimal Basis: Example (cont.)

40

3rd Step

- H doesn't change as all LHS in H are single attributes

4th Step

- H doesn't change

Minimal Basis: M = H = {A->C, B->A, D->B}

40

Lecture Example

41

41

Lecture Example

42

42

A Geometric View of FDs

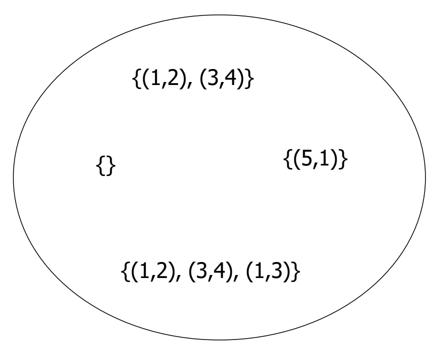
43

- Imagine the set of all *instances* of a particular relation.
- That is, all finite sets of tuples that have the proper number of components.
- Each instance is a point in this space.

43

Example: R(A,B)

44



44

An FD is a Subset of Instances

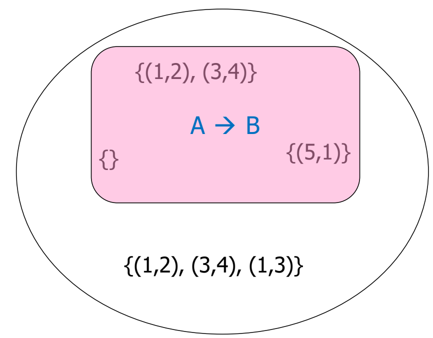
45

- For each FD $X \rightarrow A$ there is a subset of all instances that satisfy the FD.
- We can represent an FD by a region in the space.

45

Example: A → B for R(A,B)

46



46

Representing *Sets* of FDs

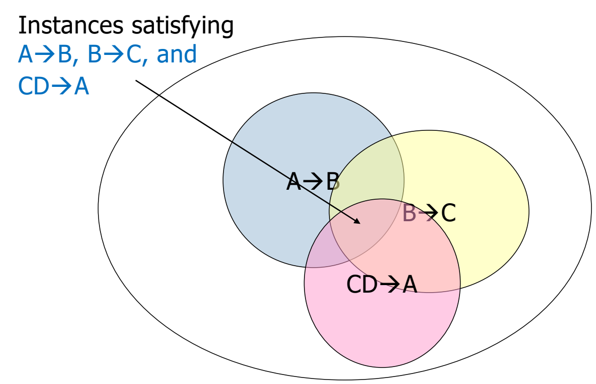
47

- If each FD is a set of relation instances, then a collection of FDs corresponds to the intersection of those sets.
  - Intersection = all instances that satisfy all of the FDs.

47

Example

48



48

Implication of FDs

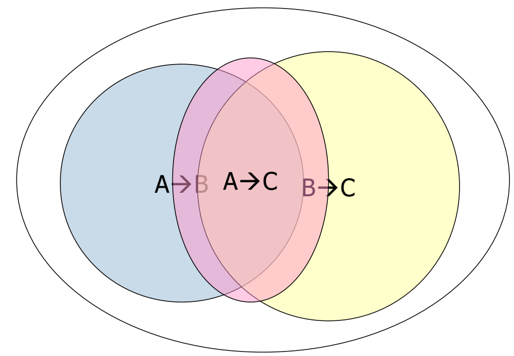
49

- If an FD $Y \rightarrow B$ follows from FDs $X_1 \rightarrow A_1, \ldots, X_n \rightarrow A_n$, then the region in the space of instances for $Y \rightarrow B$ must include the intersection of the regions for the FDs $X_i \rightarrow A_i$.
  - That is, every instance satisfying all the FDs $X_i \rightarrow A_i$ surely satisfies $Y \rightarrow B$.
  - But an instance could satisfy $Y \rightarrow B$, yet not be in this intersection.
- For a set of FDs ***F***, ***F***$^+$ (the closure of ***F***) is the set of all FDs implied by ***F***

49

Example

50



50

Closure of F

51

- For a set of FDs ***F***, ***F***$^+$ (the closure of ***F***) is the set of all FDs that can be derived (implied) from ***F***
  - Do not confuse closure of F with closure of an attribute set

51

Closure of F

52

- Example: Assume R(A, B, C, D), with F = {A→ B, B→C). Then F$^+$ includes the following FDs:

$$A \to A,\ A \to B,\ A \to C,\ B \to B,\ B \to C,\ C \to C,\ D \to D,$$
$$AB \to A,\ AB \to B,\ AB \to C,\ AC \to A,\ AC \to B,\ AC \to C,$$
$$AD \to A,\ AD \to B,\ AD \to C,\ AD \to D,\ BC \to B,\ BC \to C,$$
$$BD \to B,\ BD \to C,\ BD \to D,\ CD \to C,\ CD \to D,$$
$$ABC \to A,\ ABC \to B,\ ABC \to C,\ ABD \to A,\ ABD \to B,$$
$$ABD \to C,\ ABD \to D,\ BCD \to B,\ BCD \to C,\ BCD \to D,$$
$$ABCD \to A,\ ABCD \to B,\ ABCD \to C,\ ABCD \to D.$$

52