Funktionale Abhängigkeiten

Übersicht

1. Motivation
2. Definitionen: Funktionale Abhängigkeiten, Schlüssel
3. Ableitungsregeln
4. Hülle von funktionalen Abhängigkeiten
5. Äquivalenz von Mengen von Abhängigkeiten
6. Minimale Überdeckung

1. Motivation

Ein Objekttyp der realen Welt entspricht dem Relationenschema R, die Objektexemplare der realen Welt entsprechen den Tupeln t(R).

Funktionale Abhängigkeiten: wichtige statische Integritätsbedingungen, deren Existenz auch Bedeutung für die Beurteilung der Qualität eines Datenbankentwurfs hat

2. Definitionen

Im folgenden bezeichnen:

V,W,X,Y,Z	Attributmengen,
A,B,C, ... L	Attribute,
Q,R,S	Relationenschemata,
q(Q), r(R), s(S)	Relationen
t,u,v,w	Tupel

Definition: Funktionale Abhängigkeiten

Seien X und Y Teilmengen von R.

Eine Relation r(R) erfüllt (satisfies) die *funktionale Abhängigkeit* (functional dependency) FD X --> Y, wenn für jedes Paar von Tupeln u,v $\in$ r(R) gilt:

$$u(X) = v(X) \text{ ----> } u(Y) = v(Y)$$

X heißt die linke Seite (left hand side) LHS,

Y heißt die rechte Seite (right hand side) RHS von FD.

Beispiel: Funktionale Abhängigkeiten

F = {PERSONALNR $\rightarrow$ ABTEILUNG, ABTEILUNG $\rightarrow$ ADRESSE}

ANG	(<u>PERSONALNR</u>	ABTEILUNG	ADRESSE)
	12	Buchhaltung	1010 Wien
	13	Verkauf	1030 Wien
	11	Filiale	4020 Linz
	14	Buchhaltung	1010 Wien

Abhängigkeiten erfüllt.

Beispiel: Funktionale Abhängigkeiten

ANG (<u>PERSONALNR</u>	ABTEILUNG	ADRESSE)
12	Buchhaltung	1010 Wien
13	Verkauf	1030 Wien
11	Filiale	4020 Linz
14	Buchhaltung	5010 Salzburg

Abhängigkeit ABTEILUNG $\rightarrow$ ADRESSE verletzt.

Definition: Schlüssel

Gegeben sei ein Relationenschema R und eine Menge F von FDs.

X $\subseteq$ R ist ein Oberschlüssel für R
genau dann, wenn X $\rightarrow$ R

X ist ein Schlüssel für R genau dann, wenn X $\rightarrow$ R und
X minimal ist

3. Ableitungsregeln

Problemstellung:

Gegeben ist eine Menge von funktionalen Abhängigkeiten F. Welche zusätzlichen Abhängigkeiten f werden durch F impliziert (F |= f) ?

Ableitungsregeln (1)

Armstrong 1974

F1: Reflexivität (reflexivity)

$Y \subseteq X \quad \vdash \quad X \rightarrow Y$

F2: Erweiterung (augmentation)

$X \rightarrow Y \quad \vdash \quad XZ \rightarrow Y$

F3: Vereinigung (additivity)

$X \rightarrow Y, X \rightarrow Z \quad \vdash \quad X \rightarrow YZ$

F4: Zerlegung (projectivity)

$X \rightarrow YZ \quad \vdash \quad X \rightarrow Y, X \rightarrow Z$

Ableitungsregeln (2)

F5: Transitivität (transitivity)

$X \rightarrow Y, Y \rightarrow Z \quad \vdash \quad X \rightarrow Z$

F6: Pseudotransivität (pseudotransitivity)

$X \rightarrow Y, YW \rightarrow Z \quad \vdash \quad XW \rightarrow Z$

Bsp.: Pseudotransitivität

$$\left.\begin{array}{l} \text{Vorl} \rightarrow \text{Leiter} \\ \text{Leiter Uhrzeit} \rightarrow \text{Hörsaal} \end{array}\right\} \vdash \quad \text{Vorl Uhrzeit} \rightarrow \text{Hörsaal}$$

Beispiel: Ableitungsregeln

F = {PERSONALNR → ABTEILUNG, ABTEILUNG → ADRESSE}

Aus F folgt auch: PERSONALNR → ADRESSE

|== ... Gültigkeitsbegriff, modelltheoretisch

|— ... Ableitbarkeitsbegriff, beweistheoretisch

Ziel: Axiomatisierung derart, daß (F |== f <===> F |— f)

Korrektheit und Vollständigkeit

Die Axiome F1 bis F6 sind *korrekt (sound)*, d.h. sie leiten von F nur wirklich gültige FD ab (F |— f ===> F |== f), und *vollständig (complete)*, d.h. sie erzeugen bei wiederholter Anwendung alle von F implizierten Abhängigkeiten (F |== f ===> F |— f)

(Beweis in Ullman oder Maier)

4. Hülle von funktionalen Abhängigkeiten

- Hülle einer Menge funktionaler Abhängigkeiten
- Hüllt einer Menge von Attributen
- Membership-Problem
- Algorithmus für das Membership-Problem

Definition: Hülle

Die Menge aller funktionalen Abhängigkeiten, die von F auf R impliziert werden, heißt die *Hülle (closure)* von F auf R und wird als $F_R{}^+$ notiert.

Ist die Attributmenge aus dem Kontext klar, wird nur F^+ geschrieben.

$$F^+ = \{f \mid F \models f\}$$

$X \rightarrow Y$ heißt *trivial*, wenn $Y \subseteq X$.

$X \rightarrow Y$ ist trivial genau dann, wenn $Y \subseteq X$

Hülle einer Menge von Attributen

Unter der Hülle einer Menge von Attributen $X \subseteq R$, geschrieben als X^+, bezüglich F versteht man die Menge aller Attribute A, sodaß $X \rightarrow A$ aus F^+ ist.

$$X^+ = \{ A \mid (X \rightarrow A) \in F^+ \}$$

$$X \equiv Y \Longleftrightarrow X^+ = Y^+$$

Beispiel: Hülle

$G = g1\ \{$ PERSONAL ABT $\rightarrow$ ORT
g2 PERSONALNR $\rightarrow$ ABT
g3 ABT $\rightarrow$ ORT $\}$

$G^+ = G \cup$ alle trivialen Abhängigkeiten $\cup$

g4 {	PERSONALNR $\rightarrow$ ORT	g2,g3
g5	PERSONALNR $\rightarrow$ ORT ABT	g2,g4
g6	PERSONALNR ABT $\rightarrow$ ORT ABT	g5
g7	PERSONALNR ORT $\rightarrow$ ABT	g2
g8	PERSONALNR ORT $\rightarrow$ ORT ABT	g5
	....	}

$\{PERSONALNR\}^+ = \{PERSONALNR, ABT, ORT\}$

Membership-Problem

Problemstellung:

Gegeben ist eine Menge funktionalen Abhängigkeiten F. Welche zusätzlichen Abhängigkeiten werden durch F impliziert ?

Ist $X \rightarrow Y \in F^+$?

Algorithmen: Closure und Member

CLOSURE (X,F)
{input: X ... Attributmenge
F ... Menge von FDs
output: X^+ }
$X^+ = X$
while $(Y \rightarrow Z) \in F$ mit $Y \subseteq X^+$ und $Z \not\subseteq X^+$ *do*
$X^+ = X^+ \cup Z$ end
return (X^+)

MEMBER $(X \rightarrow Y,F)$
{input: $X \rightarrow Y$ ... FDs
F ... Menge von FDs
output: TRUE, if $F \models X \rightarrow Y$
FALSE, if $F \not\models X \rightarrow Y$ }
return ($Y \subseteq$ CLOSURE (X,F))

Algorithmen

Anmerkung:

Der Algorithmus CLOSURE kann wie folgt effizient implementiert werden (Grundidee):

- führe für jedes Attribut eine Referenzliste, die angibt, in welchen FDs aus F dieses Attribut vorkommt (ATTRLIST)
- führe zu jeder FD einen Zähler, der angibt, wieviele Attribute der linken Seite noch nicht in X^+ sind (ZÄHLER)

Algorithmen

- wenn der ZÄHLER für eine FD Null wird, wissen wir, daß die linke Seite von FD im augenblicklichen X^+ enthalten ist, und wir die Attribute der rechten Seite von FD zu X^+ hinzufügen können: Dabei muß für jedes neue Attribut A in X^+, dessen ATTRLIST - die ja angibt, in welchen FDs A in der linken Seite vorkommt, - durchgegangen werden. In diesen FDs muß nun der ZÄHLER um 1 erniedrigt werden, da A in X^+ neu aufgenommen wurde.

Beispiel

Gegeben: F= $\{A \rightarrow B, B \rightarrow DE, D \rightarrow F, BF \rightarrow C\}$ und $A \rightarrow C$

Gilt $A \rightarrow C \in F^+$?

X^+ in CLOSURE nimmt folgende Werte an:

{A} (wegen $A \rightarrow C$ als zu überprüfende FA)

{A, B} (wegen $A \rightarrow B$)

{A, B, D, E} (wegen $B \rightarrow DE$)

{A, B, D, E, F} (wegen $D \rightarrow F$)

{A, B, C, D, E, F} (wegen $BF \rightarrow C$)

C ist Element von X^+, daher gilt $A \rightarrow C \in F$

Beispiel

Gegeben sei F= $\{A \rightarrow B, C \rightarrow DE, E \rightarrow B\}$.

Frage: Gilt $A \rightarrow C \in F^+$?

Nein, da C in keiner FA in F auf der rechten Seite vorkommt. Daher kann der CLOSURE-Algorithmus ausgehend von {A} nie das Element C in X^+ aufnehmen.

5. Äquivalenz v. Mengen v. Abhängigkeiten

Definition:

Zwei Mengen funktionaler Abhängigkeiten G und F sind *äquivalent*, geschrieben als $G \equiv F$, wenn $F^+ = G^+$ gilt.

F heißt dann *Überdeckung (cover)* von G.

Algorithmus: $F \equiv G$?

(a) prüfe für alle $g \in G$, ob $F \models g$,

(b) prüfe für alle $f \in F$, ob $G \models f$.

Beispiel:

$F = \{$PERSONALNR ABT $\rightarrow$ ORT
PERSONALNR $\rightarrow$ ABT
ABT $\rightarrow$ ORT
PERSONALNR $\rightarrow$ ORT$\}$
$G = \{$PERSONALNR ABT $\rightarrow$ ORT
PERSONALNR $\rightarrow$ ABT
ABT $\rightarrow$ ORT$\}$

Frage: Ist $F \equiv G$?
(a) prüfe für alle $g \in G$, ob $F \models g$ erfüllt
(b) prüfe für alle $f \in F$, ob $G \models f$ erfüllt

Frage: $G \models \{$PERSONALNR $\rightarrow$ ORT$\}$?
Ja, folgt aus der Transitivität von g2 und g3 $\rightarrow F \equiv G$.

6. Definition: Minimale Überdeckung (1)

Eine Menge funktionaler Abhängigkeiten F heißt *minimal*, wenn:

1. F ist *kanonisch*: für jede FD $\in$ F besteht die rechte Seite nur aus einem Attribut

2. F ist *nichtredundant*:
   für keine FD $X \rightarrow A \in F$ gilt $(F \setminus \{X \rightarrow A\}) \equiv F$

Definition: Minimale Überdeckung (2)

3. Keine FD $\in$ F enthält überflüssige (extraneous) Attribute auf der linken Seite: für keine FD $X \rightarrow A \in F$ und echte Teilmenge Z von X gilt $(F \setminus \{X \rightarrow A\} \cup \{Z \rightarrow A\}) \equiv F$

Theorem:

Zu jeder Menge funktionaler Abhängigkeiten F gibt es eine minimale Überdeckung.

(Beweis Ullman)

Algorithmen: Minimale Überdeckung

(a) Test auf redundante Abhängigkeiten (ist $X \rightarrow Y$ in F redundant) :

$X \rightarrow Y \in (F \setminus \{X \rightarrow Y\})^+$?

(b) Test auf überflüssige Attribute (ist B in $XB \rightarrow Y$ überflüssig) :

$X \rightarrow Y \in F^+$?

Beispiel: Minimale Überdeckung (1)

G = {PERSONALNR → ABT ORT
ABT → ORT
PERSONALNR ABT →NAME}

kanonisch:

G' = {PERSONALNR → ABT
PERSONALNR → ORT
ABT → ORT
PERSONALNR ABT → NAME}

Beispiel: Minimale Überdeckung (2)

nichtredundant:

Frage:

Ist PERSONALNR → ORT redundant in G' ?

Prüfe, ob aus G' \ {PERSONALNR → ORT} =

G" = {PERSONALNR → ABT g1"
ABT → ORT g2"
PERSONALNR ABT →NAME}

folgt, daß {PERSONALNR → ORT} gilt.

Antwort: Ja, aufgrund der Transitivität von g1", g2".

Beispiel: Minimale Überdeckung (3)

keine überflüssigen Attribute:

Frage: Ist ABT in PERSONALNR ABT → NAME in G" überflüssig ?

Prüfe, ob aus G" folgt: PERSONALNR → NAME

Antwort: Ja.

$\{\text{PERSONALNR}\}^+$ = {PERSONALNR, ABT, NAME, ORT}

Daher gilt: PERSONALNR → NAME

Die *minimale Überdeckung* für G ist somit:

G" = {PERSONALNR → ABT
ABT → ORT
PERSONALNR → NAME}

Beispiel: Funktionale Abhängigkeiten

Gegeben ist die Relation r(R):

r	(A	B	C	D	E)
	a1	b1	c1	d1	e1
	a1	b2	c2	d2	e1
	a2	b1	c3	d3	e1
	a2	b1	c4	d3	e1
	a3	b2	c5	d1	e1

Geben Sie an, welche der folgenden Abhängigkeiten r erfüllt

(a) A --> D (b) AB --> D (c) C --> BDE

(d) E --> A (e) A --> E (f) A -->> BC

Beispiel

F = {AB → C, DE → AB, A → D, B → E, DE → C, AD → G}

Geben Sie F in kanonischer Darstellung an.

(AB)+ = ?

(DE)+ = ?

Gilt AB ≡ DE ?

Gilt AB → CDG ?

Ist DE → C in F redundant ? (Warum, warum nicht ?)

Ist D in AD → G überflüssig ? (Warum, warum nicht ?)

Geben Sie eine minimale Überdeckung von F an.

Lösung (1)

$(AB)^+$ = ABCDEG

$(DE)^+$ = ABCDEG

AB ≡ DE, weil $(AB)^+ = (DE)^+$

AB → CDG gilt, weil CDG in $(AB)^+$

DE → C redundant, weil FA aus {AB → C, DE → AB} aufgrund der Transivität folgt

D in AD → G überflüssig wegen A → D

Lösung (2)

Kanonische Form:
F={AB → C, DE → A, DE → B, A → D, B → E,
DE → C, AD → G}

Redundante FA: AB → C wegen A → D, B → E, DE → C

Überflüssige Attribute:
D in AD → G wegen A → D

Minimale Überdeckung
F‘= {DE → A, DE → B, A → D, B → E, DE → C, A → G}

Inclusion Dependencies

“Inklusionsabhängigkeiten”, IA’s: Teilmengenbeziehungen
R[X] ⊆ S[Y] *|X| = |Y|*

r(R) und *s(S)* erfüllen die ID *R[X]* ⊆ S[Y] falls $\pi_x(r) \subseteq \pi_y(s)$

Bsp.: personen (PNR ZUNAME VORNAME GEHALT ...)
kinder (PNR VORNAME ...)

kinder[PNR] ⊆ personen [PNR]

Ableitungsregeln für IAen

Reflexivität:

$X \subseteq R \vdash R[X] \subseteq R[X]$

Projektion und Permutation:

$R[X_1 \ldots X_n] \subseteq S[Y_1 \ldots Y_n] \vdash R[X_k \ldots X_m] \subseteq S[Y_k \ldots Y_m]$,
wobei $k \ldots m$ Sequenzen von Indizes aus dem Bereich $1 \ldots n$ sind

Transitivität:

$R[X] \subseteq S[Y], S[Y] \subseteq T[Z] \vdash R[X] \subseteq T[Z]$