

Fuzzy Funktionale Abhängigkeiten (FFD)

Julie Orzea

Betreuer: Heiko Schepperle

Fuzzy Funktionale Abhängigkeiten (FFD)

- Fuzzy Datenbanken
- Verschiedene (fuzzy) Funktionale Abhängigkeiten
- Inferenz-Regeln, Schlüssel und q-Schlüssel und andere Erweiterungen
- Verwendung
 - Zerlegung in Normalformen
 - Entscheidungshilfe
- Zusammenfassung

Modellierung von Imperfektion

- Wird verwendet, um besser die Realität zu modellieren
- 3 verschiedenen Arten
 - Ein Attribut, das nicht sicher ist
 - {Carine; (Französisch/0,3)}
 - Ein Attribut mit mehreren Werten
 - {Carine;(Französisch/0,8; Deutsch/0,2)}
 - Ein Attribut mit Unsicherheit
 - {Carine, jung}

Funktionale Abhängigkeiten

- Eine funktionale Abhängigkeit:

$A \rightarrow B$

$$\Leftrightarrow \forall (t, t') \in R, t.A = t'.A \Rightarrow t.B = t'.B$$

- Wie kann man die Fuzziness in funktionale Abhängigkeiten einführen?

Problem, wenn in den Attributen
Unsicherheit ist

Fuzzy Funktionale Abhängigkeiten und die Ähnlichkeits-Funktion(en)

Man will folgendes benutzen:

Wenn $t.A$ und $t'.A$ ähnlich sind, müssen auch $t.B$ und $t'.B$ ähnlich sein:

$$A \rightarrow B \iff \forall (t, t') \in R, t.A \approx t'.A \Rightarrow t.B \approx t'.B$$

„Ähnlich“ oder „Resemblance“ (\approx) muss durch eine Funktion $RES_X : [\text{dom}(X)][\text{dom}(X)] \rightarrow [0, 1]$ auf A (RES_A) und eine auf B (RES_B) definieren werden.

Fuzzy Implikations-Funktion

Der allgemeine Fall:

$A \sim_{\theta} B$ ist gültig auf $R(U) \Leftrightarrow$

$$\left(\min_{t, t' \in R} I[t.A \approx t'.A; t.B \approx t'.B] \right) \geq \theta$$

wobei I eine Fuzzy Implikations-Funktion ist,

$$I : [0, 1] [0, 1] \rightarrow [0, 1]$$

"minimale Wahrscheinlichkeit, dass $A \approx A'$ und $B \approx B'$ gilt"

Implikations-Funktion in den nicht Fuzzy Fall

Der allgemeine Fall:

θ

$\sim \rightarrow_{\theta}$

\approx

$I[X, Y]$

Der nicht-fuzzy Fall:

1

\rightarrow

=

$X \Rightarrow Y$ (true=1, false=0)

$A \sim \rightarrow_{\theta} B$ ist gültig auf $R(U)$

\Leftrightarrow

$$(\min_{t, t' \in R} I[t.A \approx t'.A; t.B \approx t'.B]) \geq \theta$$

$A \rightarrow B$ ist gültig auf $R(U)$

\Leftrightarrow

$$(\min_{t, t' \in R} I[t.A = t'.A; t.B = t'.B]) \geq 1$$

$$(t.A = t'.A) \Rightarrow (t.B = t'.B)$$

Fuzzy Funktionale Abhängigkeiten

- **Fuzzy Funktionale Abhängigkeiten von Raju und Majumdar:**

$A \sim_{rg} B$ ist gültig auf $R(U)$

$$\Leftrightarrow \forall (t, t') \in R, RES_A(t.A, t'.A) \Rightarrow_{rg} RES_B(t.B, t'.B)$$
$$a \Rightarrow_{rg} b = 1, \text{ wenn } a \leq b; \quad 0, \text{ sonst}$$

- **Fuzzy Funktionale Abhängigkeiten von G. Chen et al:**

$A \sim_{\lambda} B$ ist gültig auf $R(U)$ mit λ in $]0;1]$

$$\Leftrightarrow \forall (t, t') \in R, RES_A(t.A, t'.A) \Rightarrow_{\lambda} RES_B(t.B, t'.B)$$
$$a \Rightarrow_{\lambda} b = 1 \text{ wenn } a \leq b; \quad b \text{ sonst}$$

- **Fuzzy Funktionale Abhängigkeiten von Cubero et al:**

$A \sim_{(\lambda, \mu)} B$ ist gültig auf $R(U)$ mit λ und μ in $]0;1]$

$$\Leftrightarrow \forall (t, t') \in R, RES_A(t.A, t'.A) \Rightarrow_{(\lambda, \mu)} RES_B(t.B, t'.B)$$
$$a \Rightarrow_{(\lambda, \mu)} b = 1 \text{ wenn } a > \lambda \text{ und } b > \mu; \quad 0 \text{ sonst}$$

Welche Funktion ?

- Alle haben Vorteile und Nachteile
 - entweder leicht nachvollziehbar, verständlich
 - oder leichte Übertragbarkeit von Regeln für FD auf FFD

Ableitungs-Regeln / Amstrong-Axiome

Amstrong-Axiome (Analog zu FD):

- Wenn Y in X , dann $X \sim_{\theta} Y$ („**Reflexivität**“)
- Wenn $X \sim_{\theta} Y$, dann $XZ \sim_{\theta} YZ$ („**Expansivität**“)
- Wenn $X \sim_{\theta} Y$ und $Y \sim_{\gamma} Z$, dann $X \sim_{\mu} Z$ mit $\mu = \min(\theta, \gamma)$ („**Transitivität**“)

Beziehungsweise gibt es die **Ableitungs-Regeln** :

- Wenn $X \sim_{\theta} Y$ und $Y \sim_{\gamma} Z$, dann $X \sim_{\mu} YZ$ mit $\mu = \min(\theta, \gamma)$
- Wenn $X \sim_{\theta} Y$ und $WY \sim_{\gamma} Z$, dann $XW \sim_{\mu} Z$ mit $\mu = \min(\theta, \gamma)$
- Wenn $X \sim_{\theta} Y$ und Z in Y , dann $X \sim_{\theta} Z$
- Wenn $X \sim_{\theta} Y$, dann $X \sim_{\gamma} Y$ mit alle $\gamma \leq \theta$
- $X \sim_{\theta} Y \Leftrightarrow X \sim_{\theta} Y_1, Y_2, \dots, Y_N \Leftrightarrow X \sim_{\theta} Y_i$ für $i=1..k$

Andere Erweiterungen

- **voll fuzzy funktional abhängig:**
Y ist von X voll fuzzy funktional abhängig mit einem Grad q ($X \sim \rightarrow_q Y$ („voll“)), wenn $X \sim \rightarrow_q Y$ und keine echte Teilmenge A von X gibt, so dass $A \neq \emptyset$ und $A \sim \rightarrow Y$.
- **partiell fuzzy funktional abhängig:**
Y ist von X partiell fuzzy funktional abhängig, wenn Y von X fuzzy funktional abhängig ist, aber nicht voll fuzzy funktional abhängig.
- **Hülle:**
 F^+ ist die Menge aller fuzzy funktionalen Abhängigkeiten, die man durch die Ableitungsregeln erzeugen kann.
- **Äquivalenz zweier Mengen** von FFDs: $F \Leftrightarrow G$, wenn $F^+ = G^+$

Schlüssel und q-Schlüssel

- Ein **Schlüsselkandidat** K in der Relation $R(U)$:
 $K \rightarrow U$ („voll funktional abhängig“)
„Dieselben Werte von K definieren dieselben Werte von U .“
- Ein **q-Schlüsselkandidat** K in der fuzzy Relation $R(U)$:
 $K \rightsquigarrow_q U$ („voll fuzzy funktional abhängig“)
„Dieselbe Werte von K definieren dieselben Werte von U und ähnliche Werte von K definieren ähnliche Werte von U .“
- Ein **Schlüssel** (bzw. ein **q-Schlüssel**) wird aus der Menge der Schlüsselkandidaten (bzw. der q-Schlüsselkandidaten) gewählt.
- Ein **Schlüssel** (bzw. ein **q-Schlüssel**) kann als Referenz in einer anderen Relation benutzt werden. In dieser Relation ist der Schlüssel Fremdschlüssel.

Verwendung dieser Funktionen

- Um Datenbanken zu strukturieren
 - Integritätsbedingungen
 - Redundanz und Anomalien vermeiden
 - => Zerlegung in Normalformen
- Als Entscheidungshilfe

1. q-Normalform

Jedes Attribut in der Relation muss einen einzigen Wert haben.

Alle Attribute sind atomar,

(können aber eine Wahrscheinlichkeit enthalten oder mehrere Werte existieren, von denen ausschließlich einer wahr ist.)

{Carine, Deutsch/0,7 und Englisch/0,7} geht nicht,

{Carine, Deutsch/0,5 oder Englisch/0,5} ok

2. q-Normalform

R ist in der 2. q-Normalform wenn:

- R in der **1. q-Normalform** ist und
- alle Attribute sind **voll** fuzzy funktional vom Schlüssel abhängen
"für den q-Schlüssel K von R und für eine Menge von Nichtschlüsselattributen A gilt: wenn $K \rightarrow_q A$ dann $K \rightarrow_q A$ (voll)."

3. q-Normalform

R ist in der 3. q-Normalform wenn:

- R in der **2. q-Normalform** ist und
- alle nicht q-Schlüsselattribute sind **direkt** fuzzy abhängig von der q-Schlüssel.

Zum Beispiel:

$R = \{P, H, S, C\}$,

$\{P, H \sim\rightarrow_q C; H, S \sim\rightarrow_q P; H, C \sim\rightarrow_q S; S \sim\rightarrow_q P\}$

q-Boyce-Codd Normalform

R ist in der q-Boyce-Codd Normalform wenn:

- R in der **3. q-Normalform** ist und
- es gibt kein $X \sim \rightarrow_q Y$ in F^+ , wo X nicht ein **q-Super-Schlüssel** ist.

Zum Beispiel:

$R = \{P, H, S, C\}$,

$\{P, H \sim \rightarrow_q C; H, S \sim \rightarrow_q P; H, C \sim \rightarrow_q S\}$

Anwendungsmöglichkeiten

$K \rightarrow_q U$:

„Dieselben Werte von K definieren dieselben Werte von U und ähnliche Werte von K definieren ähnliche Werte von U.“

Man kann damit:

- fuzzy **Integritätsbedingungen** beschreiben,
- unbekannte Attribute **extrapolieren/interpolieren**,
- fuzzy Datebanken für **logisches Schließen verwenden** ("analogical reasoning and cooperative answering"),
- die Realität besser als **fuzzy Mini-Welt** darstellen.

Zusammenfassung

- Wie kann man die Datenbanktheorie zu fuzzy Datenbanken erweitern?
 - Ähnlichkeitsregeln
 - fuzzy funktionale Abhängigkeiten
 - fuzzy Ableitungsregeln
 - fuzzy Schlüssel
- Anwendungsmöglichkeiten:
 - besser die Realität beschreiben
 - Datenbanken strukturieren (Normalformen)
 - Integritätsbedingungen beschreiben
 - unbekannte Attribute extra- oder interpolieren
- Um etwas mit diesen Datenbanken und Relationen zu entscheiden, muss man bei der Festlegung der Erweiterungen und der Ähnlichkeitsfunktion sehr aufpassen.