Συναρτησιακές Εξαρτήσεις & Κανονικές Μορφές

ΗΥ - 360 Αρχεία και Βάσεις Δεδομένων

6ο Φροντιστήριο

1

Συναρτησιακές Εξαρτήσεις

- Μια συναρτησιακή εξάρτηση μεταξύ X και Y συμβολίζεται με $X \rightarrow Y$ και ορίζει έναν περιορισμό ως προς τις τιμές των χαρακτηριστικών. Σε αυτή την περίπτωση το σύνολο χαρακτηριστικών Y λέγεται **συναρτησιακά εξαρτώμενο** (*functional dependent*) από το σύνολο X.
- Αν δύο πλειάδες έχουν τις ίδιες τιμές στα χαρακτηριστικά του συνόλου X, τότε θα έχουν ίδιες τιμές και στα χαρακτηριστικά του συνόλου Y.
- Σε οποιαδήποτε σχέση R, αν $S_1 \subseteq R$ και $S_2 \subseteq R$, η συναρτησιακή εξάρτηση $S_1 \rightarrow S_2$ καθορίζει ότι για οποιεσδήποτε δύο πλειάδες t_1 και t_2 της R ισχύει

  $$\text{Αν } t_1[S_1] = t_2[S_1], \text{ τότε } t_1[S_2] = t_2[S_2]$$

  ,δηλαδή αν οποιεσδήποτε δυο πλειάδες μιας σχέσης της R συμφωνούν (έχουν την ίδια τιμή) σε κάποια γνωρίσματα $S_1 \subseteq R$ τότε συμφωνούν (έχουν την ίδια τιμή) και σε κάποια γνωρίσματα $S_2 \subseteq R$.
- Σε ένα σύνολο χαρακτηριστικών μπορεί να υπάρχουν **τετριμμένες** (trivial) συναρτησιακές εξαρτήσεις αλλά και άλλες που μπορούν να συναχθούν.

$\pi.\chi.$

1) Έστω μια αλυσίδα καταστημάτων. Κάθε τμήμα έχει έναν διευθυντή και κάθε διευθυντής δεν μπορεί να ανήκει σε παραπάνω από ένα τμήμα. Ένα τμήμα μπορεί σε πολλές πόλεις.

- Ισχύει η συναρτησιακή εξάρτηση,

$$SupSSN \rightarrow DeptNo$$

- Δεν ισχύουν

$$SupSSN \rightarrow City$$

$$DeptNo \rightarrow City$$

#	DeptNo	SupSSN	City	#
	2	1234	N.York	
	3	5432	Paris	
	2	1234	Barcelona	
	3	5432	Paris	
	2	1234	Chicago	

2) Έστω ότι ένας διευθυντής διευθύνει μόνο ένα τμήμα σε κάθε πόλη.

- Ισχύει η συναρτησιακή εξάρτηση,

$$\{City, DeptNo\} \rightarrow SupSSN$$

- Δεν ισχύουν

$$SupSSN \rightarrow \{City, DeptNo\}$$

$$\{SupSSN, DeptNo\} \rightarrow City$$

#	DeptNo	SupSSN	City	#
	2	1234	N.York	
	3	5432	Paris	
	2	1234	Barcelona	
	3	5432	Paris	
	2	1234	Chicago	

Αξιώματα Armstrong

Είναι **βάσιμα** καθώς δεν δίνουν λανθασμένες εξαρτήσεις και **πλήρη** καθώς δίνουν όλες τις συναγόμενες συναρτησιακές εξαρτήσεις.

- Ανακλαστικότητα (Reflexivity):

$$\text{Αν } B \subseteq A, \text{ τότε } A \rightarrow B$$

- Προσαύξηση (Augmentation):

$$\text{Αν } A \rightarrow B, \text{ τότε } AC \rightarrow BC$$

- Μεταβατικότητα (Transitivity):

$$\text{Αν } A \rightarrow B \text{ και } B \rightarrow C, \text{ τότε } A \rightarrow C$$

Συνέπειες αξιωμάτων

- Ένωση (Union):

$$\text{Αν } A \rightarrow B \text{ και } A \rightarrow C, \text{ τότε } A \rightarrow BC$$

- Διάσπαση (Decomposition):

$$\text{Αν } A \rightarrow CB, \text{ τότε } A \rightarrow B \text{ και } A \rightarrow C$$

- Ψευδομεταβατικός κανόνας (Pseudo-transitivity):

$$\text{Αν } A \rightarrow B \text{ και } BC \rightarrow D, \text{ τότε } AC \rightarrow D$$

π.χ.

1) Έστω $R = (A, B, C, G, H, I)$ και οι συναρτησιακές εξαρτήσεις

$$A \to B, \quad A \to C, \quad CG \to H, \quad CG \to I, \quad B \to H$$

Μέσω των κανόνων Armstrong εξάγονται οι,

$$A \to H, \quad A \to BC, \quad AG \to I, \quad CG \to HI$$

□

2) Δίνεται πίνακας R με τα χαρακτηριστικά W, U, V, X, Y, Z και οι συναρτησιακές εξαρτήσεις:

$$W \to UV, \quad U \to Y, \quad VX \to YZ$$

Ζητείται να αποδειχθεί ότι ισχύει $WX \to Z$.

Λύση

- Με διάσπαση από τη $W \to UV$ προκύπτει $W \to V$.
- Με επαύξηση προκύπτει $WX \to VX$.
- Με μεταβατικότητα προκύπτει $WX \to YZ$.
- Με διάσπαση προκύπτει $WX \to Z$.

□

Κλειδιά

Ορισμός

*Το K είναι **υποψήφιο κλειδί** της σχέσης R ανν $K \to R$ για κανένα $\alpha \subset K$, $\alpha \to R$.*

π.χ. Έστω σχεσιακό σχήμα με $R = (A, B, C, D, E)$ με ΣΕ:
$\{A \to BC, CD \to E, B \to D, E \to A\}$ βρίσκουμε ότι

$$CD \to ABCDE, \quad A \to ABCDE, \quad E \to ABCDE$$

Τα υποψήφια κλειδιά για την σχέση R είναι

$$A, \quad E, \quad CD$$

Δηλαδή,

- ◇ Υποψήφιο κλειδί K: υπερκλειδί με την ιδιότητα ότι αν αφαιρεθεί ένα οποιοδήποτε γνώρισμα A από το K, το K' που προκύπτει δεν είναι υπερκλειδί.
- Κάθε σχέση έχει τουλάχιστον ένα υπερκλειδί.
- Από τον ορισμό, κάθε(σχήμα) σχέσης έχει τουλάχιστον ένα (πρωτεύον) κλειδί – δεν υπάρχουν ασθενείς σχέσεις.

Κλειστότητα

Ορισμός

*Το σύνολο των συναρτησιακών εξαρτήσεων που μπορούν να παραχθούν από ένα σύνολο εξαρτήσεων F λέγεται **κλειστότητα** (closure) και συμβολίζεται με F^+.*

π.χ. Έστω $R = \{A, B, C, D\}$ και ισχύει

$$A \to B, \quad A \to C \quad B \to D$$

Αν έχουμε μια συναρτησιακή εξάρτηση όπου $A \to D$ όπου D είναι ένα υποσύνολο του A τότε η συγκεκριμένη συναρτησιακή εξάρτηση είναι τετριμμένη

$$A \to D, \quad A \to A(\text{trivial})$$

□

◇ Ο προσδιορισμός του συνόλου κλειστότητας F^+ είναι δαπανηρός αλγοριθμικά ακόμη και για μικρό σύνολο εξαρτήσεων F.

Ορισμός

Κλειστότητα του συνόλου χαρακτηριστικών *X (attribute closure), ως προς το σύνολο των χαρακτηριστικών εξαρτήσεων F, είναι το σύνολο των χαρακτηριστικών που είναι συναρτησιακά εξαρτώμενα από το σύνολο των χαρακτηριστικών X και το συμβολίζουμε με X^+.*

π.χ. Στη σχέση $\mathrm{R} = (\text{ΑΦΜ, ΑΜ, βαθμός, όνομα, διεύθυνση})$ με (ΣΕ)

$$\text{ΑΦΜ, ΑΜ} \to \text{βαθμός}, \quad \text{ΑΦΜ} \to \text{όνομα, διεύθυνση}$$

Παίρνουμε ότι $\{\text{ΑΦΜ}\}^+ = \text{ΑΦΜ, όνομα, διεύθυνση}.$ □

Σημείωση

Αν το X^+ είναι το σύνολο όλων των γνωρισμάτων του πίνακα τότε το X είναι υποψήφιο κλειδί.

Υπολογισμός κλειστότητας A^+

Αλγόριθμος υπολογισμού A^+ της κλειστότητας του A κάτω από το σύνολο F

```
αποτέλεσμα := A
while (υπάρχουν αλλαγές στο αποτέλεσμα) do
    for each β → γ στο F do
        begin
          if β ⊆ αποτέλεσμα then
            αποτέλεσμα := αποτέλεσμα ∪ γ
        end
```

Παραδείγματα

1) Έστω $R = \{A, B, C, G, H, I\}$ και οι συναρτησιακές εξαρτήσεις του F:

$$I \to B, \quad A \to H, \quad B \to G, \quad C \to G, \quad CG \to I$$

Να υπολογιστεί το $\{A, C\}^+$.

1. $X = AC$
2. $X = ACH$ από $A \to H$.
3. $X = ACGH$ από $C \to G$.
4. $X = ACGHI$ από $CG \to I$.
5. $X = ABCGHI$ από $I \to B$.

2) Έστω $R = \{A, B, C, D, E, G\}$ και οι συναρτησιακές εξαρτήσεις του F:

$$A \rightarrow B, \quad A \rightarrow C, \quad BC \rightarrow D, \quad D \rightarrow E, \quad B \rightarrow G$$

Να υπολογιστεί την A^+

1. $X = A$
2. $X = AB$ από $A \rightarrow B$.
3. $X = ABC$ από $A \rightarrow C$.
4. $X = ABCD$ από $BC \rightarrow D$.
5. $X = ABCDE$ από $D \rightarrow E$.
6. $X = ABCDEG$ από $B \rightarrow G$.

3) Δίνεται ο πίνακας $R = \{V, Y, Z, W\}$ και το

$$F^+ = \{V \rightarrow Z, VZ \rightarrow W, W \rightarrow Y, VY \rightarrow W\}$$

Να βρεθεί η κλειστότητα του χαρακτηριστικού V.

1. $X = V$
2. $X = VZ$ από $V \rightarrow Z$.
3. $X = VZW$ από $VZ \rightarrow W$.
4. $X = VXWY$ από $W \rightarrow Y$.

Ελάχιστη Κάλυψη

Έστω δύο σύνολα (ΣΕ) F, G τότε,

1. Είναι ισοδύναμα αν κάθε (ΣΕ) στο F είναι δυνατόν να παραχθεί από το G και κάθε (ΣΕ) στο G είναι δυνατόν να παραχθεί από το F, δηλαδή τα F και G είναι ισοδύναμα αν ισχύει: $F^+ = G^+$.

2. Το F καλύπτει το G αν κάθε (ΣΕ) στο G είναι δυνατόν να παραχθεί από το F, δηλαδή ισχύει: $G^+ \subseteq F^+$.

3. Ένα σύνολο F από (ΣΕ) είναι **ελάχιστη κάλυψη** αν ικανοποιεί τα παρακάτω:
    3.1 Κάθε εξάρτηση στο F είναι της μορφής: $X \to A$, όπου A είναι ένα απλό γνώρισμα.

    3.2 Δεν μπορούμε να αποσύρουμε μια (ΣΕ) από το F και να εξακολουθούμε να έχουμε ένα σύνολο ισοδύναμο του F.

    3.3 Δεν μπορούμε να αντικαταστήσουμε μια (ΣΕ) $X \to A$ από το F με μια (ΣΕ) $Y \to A$, όπου $Y \subset X$ και να εξακολουθούμε να έχουμε ένα σύνολο ισοδύναμο του F.

Αλγόριθμος υπολογισμού της ελάχιστης κάλυψης ενός συνόλου ΣΕ

- (**Διάσπαση**) Δημιουργούμε ένα νέο σύνολο G ισοδύναμο του F όπου φροντίζουμε να έχουμε (ΣΕ) με μόνο ένα γνώρισμα στο δεξιό μέλος της συνάρτησης.
- Αντικαθιστούμε τις (ΣΕ) με άλλες, που έχουν λιγότερα γνωρίσματα στο αριστερό μέλος εφόσον δεν επηρεάζουν την κλειστότητα του G
- (**Πλεονάζουσες**) Αφαιρούμε από το G όλες τις (ΣΕ) που δεν επηρεάζουν στην κλειστότητα του G αν αφαιρεθούν.
- Συγχωνεύουμε τις (ΣΕ) που έχουν το ίδιο αριστερό μέλος, φροντίζοντας έτσι όλα τα αριστερά μέλη να είναι μοναδικά.

Παράδειγμα

Έστω $R = \{A, B, C\}$ και οι συναρτησιακές εξαρτήσεις του F:

$$A \to BC, \quad B \to C, \quad A \to B, \quad AB \to C$$

Να υπολογιστεί η ελάχιστη κάλυψη του F.

Έχουμε $G = \{A \to BC, B \to C, A \to B, AB \to C\}$.

1. Η $A \to BC$ έχει πάνω από ένα γνωρίσματα στο δεξί μέλος, οπότε τη διασπάμε σε $A \to B$ και $A \to C$. Άρα,
   $G = \{A \to B, A \to C, B \to C, A \to B, AB \to C\}$.
2. Η $AB \to C$ έχει πάνω από ένα γνωρίσματα στο αριστερό μέλος. Θα κοιτάξω τις κλειστότ των A και B ως προς το $(G - (AB \to C))$.
   2.1 $A^+ = ABC$, στην οποία περιέχεται το B. Άρα μπορούμε να το βγάλουμε από την (ΣΕ).
   2.2 $G = \{A \to B, A \to C, B \to C, A \to B, A \to C\}$.
3.
   3.1 Οι δύο τελευταίες είναι επεναλαμβανόμενες, οπότε θα αφαιρεθούν.
   3.2 Λόγω μεταβατικότητας ($A \to B$, $B \to C$) η $A \to C$ καλύπτεται και θα αφαιρεθεί.
   3.3 $G = \{A \to B, B \to C\}$.

4. Δεν υπάρχουν (ΣΕ) με ίδιο αριστερό μέλος για να συγχωνέψουμε, επομένως δεν μεταβάλλεται το G.

Συνεπώς, έχουμε την ελάχιστη κάλυψη $G = \{A \to B, B \to C\}$.

Υποψήφια κλειδιά

Εύρεση υποψήφιων κλειδιών δεδομένου ενός πίνακα R και των εξαρτήσεων F μεταξύ των πεδίων του.

Βήμα 1ο: Βρες τα χαρακτηριστικά που δεν εμφανίζονται ούτε στην αριστερή αλλά ούτε και στην δεξιά πλευρά της συναρτησιακής εξάρτησης.

Βήμα 2ο: Βρες τα χαρακτηριστικά που εμφανίζονται μόνο στη δεξιά πλευρά.

Βήμα 3ο: Βρες τα χαρακτηριστικά που εμφανίζονται μόνο στην αριστερή πλευρά.

Βήμα 4ο: Συνδύασε τα χαρακτηριστικά του βήματος 1 & 3.

Βήμα 5ο: Βρες την κλειστότητα των χαρακτηριστικών του βήματος 4. Αν σε αυτήν ανήκουν όλα τα υπόλοιπα χαρακτηριστικά τότε είναι υποψήφιο κλειδί. Διαφορετικά, επανέλαβε τα βήματα 4 & 5 προσαρτώντας κάθε φορά ένα χαρακτηριστικό που δεν συμπεριλαμβάνεται στο βήμα 2 & 4.

Παράδειγμα

Έστω $R = \{A, B, C, D, E, F, G\}$ μια σχέση και $F = \{AB \rightarrow F, AD \rightarrow E, F \rightarrow G\}$ ένα σύνολο εξαρτήσεων στη σχέση αυτή. Βρείτε ένα υποψήφιο κλειδί στην R.

Βήμα 1ο: C.

Βήμα 2ο: EG.

Βήμα 3ο: ABD.

Βήμα 4ο: $ABCD$.

Η κλειστότητα του συνόλου χαρακτηριστικών του βήματος 4 είναι το $ABCDFEG$. Άρα το $ABCD$ είναι υποψήφιο κλειδί.

Lossless-Join Decomposition

Ορισμός

*Μια αποσύνθεση $\{R_1, R_2, \ldots, R_k\}$ της σχέσης R με συναρτησιακές εξαρτήσεις F λέγεται αποσύνθεση **χωρίς απώλεια πληροφορίας** (Lossless - join decomposition), αν ανεξάρτητα από το περιεχόμενο της R, οι συναρτησιακές εξαρτήσεις εξασφαλίζουν ότι*

$$R = R_1 \bowtie R_2 \bowtie \ldots \bowtie R_k$$

Θεώρημα

Δεδομένης μιας σχέσης R και ενός συνόλου (ΣΕ) F οι οποίες πληρούνται στην R μια αποσύνθεση της R στις σχέσεις $R1$ και $R2$ δεν πάσχει από απώλεια πληροφορίας αν ισχύει τουλάχιστον μία από τις δύο σχεσιακές εξαρτήσεις:

$$Head(R_1) \cap Head(R_2) \rightarrow Head(R_1)$$
$$Head(R_1) \cap Head(R_2) \rightarrow Head(R_2)$$

Παράδειγμα

π.χ. Έστω $R = \{\text{Τίτλος}, \text{Έτος}, \text{Διάρκεια}, \text{Είδος}, \text{Όνομα Ηθοποιού}, \text{Διεύθυνση}, \text{Έτος Γέννησης}\}$.

Έχουμε τις (ΣΕ)

{Τίτλος Έτος $\rightarrow$ Διάρκεια, Τίτλος Έτος $\rightarrow$ Είδος},

{Όνομα Ηθοποιού $\rightarrow$ Διεύθυνση, Όνομα Ηθοποιού $\rightarrow$ Έτος Γέννησης}

και τις σχέσεις

$R_1 = \{\text{Τίτλος}, \text{Έτος}, \text{Διάρκεια}, \text{Είδος}\}$

$R_2 = \{\text{Τίτλος}, \text{Έτος}, \text{Όνομα Ηθοποιού}, \text{Διεύθυνση}, \text{Έτος Γέννησης}\}$

Άρα $R_1 \cap R_2 = \{\text{Τίτλος}, \text{Έτος}\}$. Γνωρίζουμε ότι ένα κλειδί K είναι υπερκλειδί μιας σχέσης R αν και μόνο αν ισχύει $K \rightarrow R$. Συνεπώς η $R_1 \cap R_2$ είναι υπερκλειδί για την R_1, άρα και αποσύνθεση χωρίς απώλειες.