

Κλείσιμο Συνόλου Γνωρισμάτων

Χαρίδημος Κονδυλάκης
Τμήμα Επιστήμης Υπολογιστών,
Πανεπιστήμιο Κρήτης

Κλείσιμο Συνόλου Γνωρισμάτων

- Ο υπολογισμός του κλεισίματος ενός συνόλου από ΣΕ μας δίνει τα σύνολα όλων των γνωρισμάτων τα οποία προσδιορίζονται συναρτησιακά από άλλα σύνολα γνωρισμάτων
- Ο υπολογισμός αυτός έχει κόστος εκθετικό ως προς το μέγεθος του αρχικού συνόλου των ΣΕ.
- Δεδομένου ενός συνόλου F από ΣΕ μιας σχέσης R και ενός συνόλου X από γνωρίσματα του σχήματος της R , το **κλείσιμο του** X , X^+ είναι το μέγιστο σύνολο γνωρισμάτων Y , για τα οποία $X \rightarrow Y \in F^+$.
- Ο υπολογισμός του κλεισίματος ενός συνόλου γνωρισμάτων έχει κόστος πολυωνυμικό ως προς το μέγεθος του συνόλου αυτού (πιο αποδοτικός από τον υπολογισμό κλεισίματος συνόλου ΣΕ)

Κλείσιμο Συνόλου Γνωρισμάτων

- Ο ακόλουθος αλγόριθμος υπολογίζει το X^+ δεδομένου ενός συνόλου ΣΕ F .

$I := 0; X[I] := X;$

Επανάλαβε

$I := I + 1;$

$X[I] := X[I - 1];$

Για κάθε $Z \rightarrow W$ στο F

Αν $Z \subseteq X[I]$ τότε

$X[I] := X[I] \cup W;$

Μέχρι $X[I] = X[I - 1];$

Επέστρεψε $X^+ = X[I];$

- Ο υπολογισμός του X^+ βασίζεται στην εφαρμογή του κανόνα συσσώρευσης.

Κλείσιμο Συνόλου Γνωρισμάτων

➤ **Παράδειγμα:** $X = B, F = \{B \rightarrow CD, AD \rightarrow E, B \rightarrow A\}$

Αρχικά: $X[0] = B$

$l=1:$ $X[1] = B$

$X[1] = BCD$ (λόγω της $B \rightarrow CD$)

$X[1] = ABCD$ (λόγω της $B \rightarrow A$)

$l=2:$ $X[2] = ABCD$

$X[2] = ABCDE$ (λόγω της $AD \rightarrow E$)

$l=3:$ $X[3] = ABCDE = X[2]$

Το loop τερματίζει, επομένως $X^+ = ABCDE$

Ελάχιστη Κάλυψη (Minimum Cover)

- Χρειαζόμαστε μια μέθοδο κατασκευής μιας κάλυψης για ένα σύνολο από ΣΕ και επιπλέον η κάλυψη πρέπει να είναι ελάχιστη.
 - Ο ακόλουθος αλγόριθμος κατασκευάζει μια ελάχιστη κάλυψη M ενός συνόλου ΣΕ.
1. Δημιουργούμε ένα ισοδύναμο σύνολο H από ΣΕ με ένα μόνο γνώρισμα στο δεξί μέλος.

$$H := \emptyset;$$

Για κάθε $X \rightarrow Y$ στο F

Για κάθε A στο Y

$$H := H \cup \{X \rightarrow A\};$$

Στο τέλος του 1^{ου} βήματος, $H \equiv F$

Ελάχιστη Κάλυψη (Minimum Cover)

2. Αφαιρούμε από το H τις ΣΕ οι οποίες αν αφαιρεθούν δεν επηρεάζουν το H^+

Για κάθε $X \rightarrow A$ στο H

$J := H - \{X \rightarrow A\};$

Υπολόγισε X^+ για το J ;

Αν $A \in X^+$ τότε

$H := J;$

Το 2^ο βήμα μετατρέπει το H σε ένα μικρότερο αλλά ισοδύναμο σύνολο.

Ελάχιστη Κάλυψη (Minimum Cover)

3. Αντικαθιστούμε ΣΕ με άλλες οι οποίες έχουν λιγότερα γνωρίσματα στο αριστερό μέλος εφόσον δεν επηρεάζεται το H^+ .

Για κάθε $X \rightarrow A$ στο H

Για κάθε $B \in X$

$Y := X - \{B\};$

$J := (H - \{X \rightarrow A\}) \cup \{Y \rightarrow A\};$

Υπολόγισε Y^+ για το J και Y^+ για το H ;

Αν $(Y^+ \text{ για το } J) = (Y^+ \text{ για το } H)$ τότε

$H := J;$

Καθώς το H^+ δεν αλλάζει, το σύνολο που προκύπτει είναι ισοδύναμο με το αρχικό.

Άλλος τρόπος να ελέγξουμε ότι το H^+ δεν αλλάζει είναι να δούμε αν το Y^+ για το J περιέχει το B ή να δούμε αν υπάρχει εξάρτηση $B \rightarrow Y$

Ελάχιστη Κάλυψη (Minimum Cover)

4. Εφαρμόζουμε τον κανόνα της ένωσης στις ΣΕ με κοινό αριστερό μέλος. Υποθέτουμε ότι όλες οι ΣΕ είναι “μη-μαρκαρισμένες” στην αρχή αυτού του βήματος

$M := \emptyset;$

Για κάθε $X \rightarrow A$ στο H

Αν είναι μαρκαρισμένη, συνέχισε;

Μαρκάρισε $X \rightarrow A$

$Y := \{A\};$

Για όλες τις υπόλοιπες $X \rightarrow B$ στο H

Μαρκάρισε $X \rightarrow B$

$Y := Y \cup \{B\};$

$M := M \cup \{X \rightarrow Y\};$

Επέστρεψε $M;$

Το αποτέλεσμα είναι η ελάχιστη κάλυψη M .

Ελάχιστη Κάλυψη (Minimum Cover)

- **Παράδειγμα:** Κατασκευάστε την ελάχιστη κάλυψη M για το σύνολο $F = \{A \rightarrow AC, B \rightarrow ABC, D \rightarrow ABC\}$
1. $H = \{A \rightarrow A, A \rightarrow C, B \rightarrow A, B \rightarrow B, B \rightarrow C, D \rightarrow A, D \rightarrow B, D \rightarrow C\}$
 2.
 - (a) $A \rightarrow A$ τετριμμένη – μπορεί να αφαιρεθεί
 - (b) $A \rightarrow C$ δε μπορεί να αφαιρεθεί καθώς δεν υπάρχει άλλη ΣΕ με A στο αριστερό μέλος
 - (c) $B \rightarrow A$ δε μπορεί να αφαιρεθεί καθώς $B^+ = BC$ για το $J = H - \{B \rightarrow A\}$
 - (d) $B \rightarrow B$ τετριμμένη = μπορεί να αφαιρεθεί
 - (e) $B \rightarrow C$ μπορεί να αφαιρεθεί καθώς $B^+ = ABC$ για το $J = H - \{B \rightarrow C\}$
 - (f) $D \rightarrow A$ μπορεί να αφαιρεθεί καθώς $D^+ = DBA$ για το $J = H - \{D \rightarrow A\}$

Ελάχιστη Κάλυψη (Minimum Cover)

2. (g) $D \rightarrow B$ δε μπορεί να αφαιρεθεί καθώς $D^+ = DC$ για το $J = H - \{D \rightarrow B\}$
(h) $D \rightarrow C$ μπορεί να αφαιρεθεί καθώς $D^+ = DBAC$ για το $J = H - \{D \rightarrow C\}$
Μετά το βήμα 2, $H = \{A \rightarrow C, B \rightarrow A, D \rightarrow B\}$
3. Τα βήματα 3 και 4 δε μεταβάλλουν το H .

$$\text{Άρα } M = \{A \rightarrow C, B \rightarrow A, D \rightarrow B\}$$

Ελάχιστη Κάλυψη (Minimum Cover)

- **Παράδειγμα:** Θεωρείστε τη σχέση $R\{A, B, C, D, E, G\}$ που περιέχει δεδομένα για εργοστάσια, όπου
 - A – id διευθυντή
 - B – όνομα εργοστασίου
 - C – id εργαζομένου
 - D – μισθός εργαζομένου
 - E – φόροι εργαζομένου
 - G – bonus εργαζομένου

Τα δεδομένα ικανοποιούν τους εξής περιορισμούς:

- Κάθε διευθυντής διευθύνει ένα εργοστάσιο, αλλά ένα εργοστάσιο μπορεί να έχει πάνω από ένα διευθυντή.
- Κάθε εργαζόμενος εργάζεται σε ένα εργοστάσιο, αλλά σε κάθε εργοστάσιο εργάζονται πάνω από ένας εργαζόμενοι.
- Οι φόροι ενός εργαζομένου καθορίζονται βάσει του μισθού του
- Το bonus ενός εργαζομένου καθορίζεται από το μισθό και την πολιτική του εργοστασίου
- ...

Ελάχιστη Κάλυψη (Minimum Cover)

- **Παράδειγμα:** Θεωρείστε τη σχέση $R\{A, B, C, D, E, G\}$ που περιέχει δεδομένα για εργοστάσια, όπου

A – id διευθυντή

B – όνομα εργοστασίου

C – id εργαζομένου

D – μισθός εργαζομένου

E – φόροι εργαζομένου

G – bonus εργαζομένου

Έστω ότι το σύνολο των εξαρτήσεων που προκύπτουν είναι:

$$\{A \rightarrow B, C \rightarrow BD, C \rightarrow G, CD \rightarrow E, BCDE \rightarrow G\}$$

Εφαρμόζουμε τον αλγόριθμο για να βρούμε την ελάχιστη κάλυψη.

Ελάχιστη Κάλυψη (Minimum Cover)

1. $H = \{A \rightarrow B, C \rightarrow B, C \rightarrow D, C \rightarrow G, CD \rightarrow E, BCDE \rightarrow G\}$
2.
 - (a) $A \rightarrow B$ δε μπορεί να αφαιρεθεί καθώς δεν υπάρχει άλλη ΣΕ με A στο αριστερό μέλος
 - (b) $C \rightarrow B$ δε μπορεί να αφαιρεθεί καθώς $C^+ = CDE$ για το $J = H - \{C \rightarrow B\}$
 - (c) $C \rightarrow D$ δε μπορεί να αφαιρεθεί καθώς $C^+ = BC$ για το $J = H - \{C \rightarrow D\}$
 - (d) $C \rightarrow G$ μπορεί να αφαιρεθεί καθώς $C^+ = BCDEG$ για το $J = H - \{C \rightarrow D\}$
 - (e) $BCDE \rightarrow G$ δε μπορεί να αφαιρεθεί καθώς $\{BCDE\}^+ = BCDE$ για το $J = H - \{BCDE \rightarrow G\}$
 - (f) $CD \rightarrow E$ δε μπορεί να αφαιρεθεί καθώς το γνώρισμα E δεν εμφανίζεται στο δεξιό μέλος άλλης εξάρτησης

Ελάχιστη Κάλυψη (Minimum Cover)

3.
 - (a) Αντικαθιστούμε τη $CD \rightarrow E$ με τη $D \rightarrow E$. Επειδή στο σύνολο μας υπάρχει η εξάρτηση $C \rightarrow D$ διατηρούμε την αντικατάσταση.
 - (b) Αντικαθιστούμε τη $BCDE \rightarrow G$ με τη $CDE \rightarrow G$. Υπολογίζουμε το $(CDE)^+ = BCDEG$ το οποίο είναι ίδιο και στο αρχικό σύνολο, άρα διατηρούμε την αντικατάσταση.
 - (c) Αντικαθιστούμε τη $CDE \rightarrow G$ με τη $DE \rightarrow G$. Υπολογίζουμε το $(DE)^+ = DEG$ το οποίο είναι ίδιο και στο αρχικό σύνολο, άρα διατηρούμε την αντικατάσταση.
 - (d) Αντικαθιστούμε τη $DE \rightarrow G$ με τη $E \rightarrow G$. Επειδή στο σύνολο μας υπάρχει η εξάρτηση $D \rightarrow E$ διατηρούμε την αντικατάσταση.

Ελάχιστη Κάλυψη (Minimum Cover)

4. Μετά τις συνενώσεις, προκύπτει η ελάχιστη κάλυψη

$$M = A \rightarrow B, C \rightarrow BD, D \rightarrow E, E \rightarrow G$$

Δηλαδή id διευθυντή \rightarrow όνομα εργοστασίου

id εργαζομένου \rightarrow όνομα εργοστασίου, μισθός εργαζομένου

μισθός εργαζομένου \rightarrow φόροι εργαζομένου

φόροι εργαζομένου \rightarrow bonus εργαζομένου

Αποσύνθεση χωρίς Απώλεια Πληροφορίας (Lossless-Join Decomposition)

- Η κανονικοποίηση σχημάτων σχεσιακών ΒΔ εξαρτάται από τη δυνατότητα αποσύνθεσης σχημάτων με «μικρότερα» αποφεύγοντας συγχρόνως ανωμαλίες ενημέρωσης.
- Δεδομένης μιας σχέσης R , μια αποσύνθεση (decomposition) της R σε k σχέσεις είναι ένα σύνολο $\{R_1, R_2, \dots, R_k\}$ τέτοιο ώστε:

$$Head(R) = \bigcup_{i=1}^k Head(R_i)$$

- Δεδομένου ενός στιγμιότυπου της R , το περιεχόμενο κάθε μιας από τις σχέσεις R_i , καθορίζεται από την προβολή των πλειάδων της R στα γνωρίσματα της R_i .

Αποσύνθεση χωρίς Απώλεια Πληροφορίας

- Μια αποσύνθεση $\{R_1, R_2, \dots, R_k\}$ 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 $R = R_1 \bowtie R_2 \bowtie \dots \bowtie R_k$
- Σε μια αποσύνθεση χωρίς απώλεια πληροφορίας, μπορούμε πάντα να ανακατασκευάσουμε την αρχική σχέση από τον υπολογισμό της συνένωσης (join) των σχέσεων που προκύπτουν από την αποσύνθεση.
- Διαφορετικά, η συνένωση των σχέσεων μπορεί να δώσει πληροφορία η οποία δεν υπήρχε στην αρχική σχέση.

Παράδειγμα: Αποσύνθεση με απώλεια πληροφορίας

R (ABC)

A	B	C
a_1	100	c_1
a_2	200	c_2
a_3	300	c_3
a_4	200	c_4

R1 (AB)

A	B
a_1	100
a_2	200
a_3	300
a_4	200

R2 (BC)

B	C
100	c_1
200	c_2
300	c_3
200	c_4

R1 \bowtie R2

A	B	C
a_1	100	c_1
a_2	200	c_2
a_2	200	c_4
a_3	300	c_3
a_4	200	c_2
a_4	200	c_4

Παράδειγμα: Αποσύνθεση με απώλεια πληροφορίας

- Εξετάζοντας μόνο τις R_1 και R_2 δεν μπορούμε να πούμε από ποια σχέση προήλθαν. Θα μπορούσαν να έχουν προέλθει και από την

A	B	C
a_1	100	c_1
a_2	200	c_2
a_2	200	c_4
a_3	300	c_3
a_4	200	c_4

- Η απώλεια πληροφορίας προήλθε από τις πλειάδες $(a_2, 200, c_4)$ και $(a_4, 200, c_2)$. Εμφανίζονται στην $R_1 \bowtie R_2$ επειδή έχουν κοινή τιμή στο γνώρισμα B αλλά όχι στην αρχική σχέση.

Παράδειγμα: Αποσύνθεση με απώλεια πληροφορίας

R (ABC)

A	B	C
a_1	100	c_1
a_2	200	c_2
a_3	300	c_3

R1 (AB)

A	B
a_1	100
a_2	200
a_3	300

R2 (BC)

B	C
100	c_1
200	c_2
300	c_3

$$R1 \bowtie R2 = R$$

- Τι θα συμβεί όμως αν το περιεχόμενο της R αλλάξει και προστεθεί μια πλειάδα με τιμή που ήδη υπάρχει στο κοινό γνώρισμα;
- Δεν μπορούμε να κρίνουμε αν μια αποσύνθεση πάσχει από απώλεια πληροφορίας εξετάζοντας μόνο το περιεχόμενο των σχέσεων.
- Χρειάζονται επιπλέον κανόνες, οι οποίοι λαμβάνουν υπόψη τους τις συναρτησιακές εξαρτήσεις.

Παράδειγμα: αποσύνθεση – συναρτησιακές εξαρτήσεις

Παράδειγμα: Έστω η εξάρτηση $B \rightarrow C$ ισχύει για τη σχέση R.

R (ABC)

A	B	C
a_1	100	c_1
a_2	200	c_2
a_3	300	c_3

R1 (AB)

A	B
a_1	100
a_2	200
a_3	300

R2 (BC)

B	C
100	c_1
200	c_2
300	c_3

Η εισαγωγή της πλειάδας $(a_4, 200, c_4)$ αποτυγχάνει γιατί ισχύει η εξάρτηση $B \rightarrow C$. Η εισαγωγή της πλειάδας $(a_4, 200, c_2)$ επιτρέπεται.

Παράδειγμα: αποσύνθεση – συναρτησιακές εξαρτήσεις

- Η ακόλουθη αποσύνθεση δεν πάσχει από απώλεια πληροφορίας.

R (ABC)

A	B	C
a_1	100	c_1
a_2	200	c_2
a_3	300	c_3
a_4	200	c_2

R1 (AB)

A	B
a_1	100
a_2	200
a_3	300
a_4	200

R2 (BC)

B	C
100	c_1
200	c_2
300	c_3

Θεώρημα: κλειδιά – συναρτησιακές εξαρτήσεις

Το ακόλουθο θεώρημα προσδιορίζει τη σχέση μεταξύ κλειδιών σχέσεων και συναρτησιακών εξαρτήσεων.

Θεώρημα: Έστω μια σχέση R και ένα σύνολο γνωρισμάτων

$X \subseteq \text{Head}(R)$. Τότε οι ακόλουθες προτάσεις είναι ισοδύναμες:

- (1) X είναι κλειδί της R ,
- (2) X προσδιορίζει συναρτησιακά όλα τα γνωρίσματα στην R .

Απόδειξη:

(1) \Rightarrow (2): αν το X είναι κλειδί, τότε δεν μπορούν να υπάρχουν στην R πλειάδες οι οποίες συμφωνούν σε όλα τα γνωρίσματα του X . Άρα, δεν είναι δυνατόν να υπάρχουν πλειάδες οι οποίες συμφωνούν στις τιμές των γνωρισμάτων X και δεν συμφωνούν στα υπόλοιπα γνωρίσματα. Εξ ορισμού η συναρτησιακή εξάρτηση $X \rightarrow \text{Head}(R)$ ισχύει.

Θεώρημα: κλειδιά – συναρτησιακές εξαρτήσεις

(2) \Rightarrow (1): αν ισχύει η $X \rightarrow Head(R)$, τότε δεν είναι δυνατόν να υπάρχουν δύο πλειάδες οι οποίες συμφωνούν στα γνωρίσματα X και δεν συμφωνούν στα υπόλοιπα γνωρίσματα.

Άρα, είτε δύο πλειάδες οι οποίες συμφωνούν στα γνωρίσματα X θα συμφωνούν και στα υπόλοιπα γνωρίσματα, είτε όλες οι πλειάδες έχουν διακεκριμένες τιμές στα γνωρίσματα του X .

Καθώς δεν επιτρέπεται να υπάρχουν επαναλαμβανόμενες πλειάδες σε μια σχέση, και οι δύο αυτές περιπτώσεις μας οδηγούν στο συμπέρασμα ότι το X είναι κλειδί της R .

Άρα, (1) \equiv (2)

Θεώρημα

Μια ικανή συνθήκη για έλεγχο της ιδιότητας της μη απώλειας πληροφορίας για μια αποσύνθεση:

- **Θεώρημα:** Δεδομένης μιας σχέσης R και ενός συνόλου ΣE F οι οποίες πληρούνται στην R , μια αποσύνθεση της R στις σχέσεις $R1$ και $R2$ δεν πάσχει από απώλεια πληροφορίας αν τουλάχιστον μια από τις ακόλουθες ΣE είναι λογική συνέπεια των ΣE στο F .

$$(1) \text{Head}(R1) \cap \text{Head}(R2) \rightarrow \text{Head}(R1)$$

$$(2) \text{Head}(R1) \cap \text{Head}(R2) \rightarrow \text{Head}(R2)$$

Παράδειγμα

- **Παράδειγμα:** Έστω ότι η ΣΕ $B \rightarrow C$ ισχύει στη σχέση $R(ABC)$. Η R αποσυντίθεται στις $R1(AB)$ και $R2(BC)$. Εξετάστε αν η αποσύνθεση πάσχει από απώλεια πληροφορίας.

$$Head(R1) \cap Head(R2) = B$$

Πρέπει να δείξουμε ότι ισχύει μια από τις ΣΕ (1) $B \rightarrow AB$ και (2) $B \rightarrow BC$. Από την $B \rightarrow C$ εξάγεται η $B \rightarrow BC$ με χρήση του κανόνα επαύξησης. Άρα, η αποσύνθεση δεν πάσχει από απώλεια πληροφορίας.

Πόρισμα: Αν $\{R1, R2\}$ είναι μια αποσύνθεση χωρίς απώλεια πληροφορίας για τη σχέση R και $\{R3, R4\}$ είναι μια αποσύνθεση χωρίς απώλεια πληροφορίας της σχέσης $R2$, τότε $\{R1, R3, R4\}$ είναι μια αποσύνθεση χωρίς απώλεια πληροφορίας της R .

Παράδειγμα

- **Παράδειγμα:** Αποσύνθεση της σχέσης emp_info

emp_info

emp_id	emp_name	...	skill_id	skill_name	skill_date	skill_lvl
--------	----------	-----	----------	------------	------------	-----------

- Συναρτησιακές εξαρτήσεις:
 1. emp_id → emp_name emp_phone dept_name
 2. dept_name → dept_phone dept_mgrname
 3. skill_id → skill_name
 4. emp_id skill_id → skill_date skill_lvl

Παράδειγμα

- Η ακόλουθη αποσύνθεση δεν πάσχει από απώλεια πληροφορίας

emps

emp_id	emp_name	emp_phone	dept_name
--------	----------	-----------	-----------

depts

dept_name	dept_phone	dept_mgrname
-----------	------------	--------------

emp-skills

emp_id	skill_id	skill_date	skill_lvl
--------	----------	------------	-----------

skills

skill_id	skill_name
----------	------------