Πεπερασμένα Αυτόματα
Finite Automata

Διάλεξη 9: Θεωρία Υπολογισμού · HY-280

Χαράλαμπος Ε. Τσουρακάκης

2. Από ε-NFA σε DFA

Πού βρισκόμαστε;

Δύο διαφορετικά αυτόματα αλλά ισοδύναμα

Δύο αυτόματα M_1, M_2 είναι **ισοδύναμα** αν αποδέχονται την ίδια γλώσσα:

$$L(M_1) = L(M_2).$$

- Στο **DFA**, η κατάσταση μετά από κάθε πρόθεμα είναι μοναδική.
- Στο ε**-NFA**, μπορεί να υπάρχουν πολλές δυνατές καταστάσεις και μεταβάσεις ε.

Ερώτημα

Μπορεί ένα ε-NFA να αναγνωρίζει γλώσσα που δεν αναγνωρίζεται από κανένα DFA;

Απάντηση

Οι ακόλουθες τάξεις γλωσσών ταυτίζονται:

- οι γλώσσες που περιγράφονται από κανονικές εκφράσεις,
- οι γλώσσες που αναγνωρίζονται από ε-NFA,
- οι γλώσσες που αναγνωρίζονται από DFA.

Σύνοψη

- **Την προηγούμενη εβδομάδα:** ε-NFA $\leftrightarrow$ κανονική έκφραση
- Ξεκάθαρα DFA $\rightarrow$ ε-NFA (γιατί;)
- **Σήμερα** ε-NFA$\rightarrow$ DFA

Θεώρημα 1.19, Sipser

Για κάθε μη ντετερμινιστικό πεπερασμένο αυτόματο M (με επιτρεπτές μεταβάσεις ε), υπάρχει **ντετερμινιστικό** πεπερασμένο αυτόματο D τέτοιο ώστε

$$L(D) = L(M).$$

- Η απόδειξη είναι **κατασκευαστική**: δίνει αλγόριθμο μετατροπής.
- Αν το M έχει n καταστάσεις, το D χρειάζεται **το πολύ** 2^n καταστάσεις.

Παρατήρηση

- Ξεκάθαρα DFA $\rightarrow$ ε-NFA επειδή κάθε DFA είναι ήδη ένα ε-NFA, χωρίς καμία μετατροπή αφού το ε-NFA είναι γενίκευση του DFA: ό,τι επιτρέπει ο ορισμός του DFA επιτρέπεται και στο ε-NFA.
- Κοινώς, κάθε μετάβαση του DFA γίνεται μονοσύνολο, και δεν προσθέτουμε καμία ε-μετάβαση.
- Στον κύκλο του Kleene αυτή η κατεύθυνση είναι η «τετριμμένη» (trivial): DFA $\subseteq$ NFA $\subseteq$ ε-NFA ως περιπτώσεις του ορισμού, και η πραγματική δουλειά βρίσκεται στις αντίστροφες μετατροπές.

Συγκρατούμε τα εξής:

Ένα DFA πρέπει να «θυμάται» ολόκληρο το σύνολο των πιθανών καταστάσεων όπου θα μπορούσε να βρίσκεται το NFA, δηλαδή ένα υποσύνολο του Q. Το NFA αντίθετα «παραλληλοποιεί» και μαντεύει, και αρκεί να πετύχει μία διαδρομή. Το πλεονέκτημα αυτό αφορά τον αριθμό καταστάσεων, όχι το ποιες γλώσσες αναγνωρίζονται.

Η κεντρική ιδέα: μία κατάσταση του DFA = ένα σύνολο

Έστω $M = (K, \Sigma, \Delta, s, F)$ και ότι, μετά την ανάγνωση μιας λέξης w, το M μπορεί να βρίσκεται στις καταστάσεις

$$\{q_0, q_2, q_3\}.$$

Το DFA **δεν επιλέγει** κάποια από αυτές: κρατά ως δική του κατάσταση ολόκληρο το σύνολο $\{q_0, q_2, q_3\}$.

Μετάβαση σε γράμμα a

Από κάθε q του τρέχοντος συνόλου ακολουθούμε **όλες** τις μεταβάσεις με ετικέτα a. Μετά προσθέτουμε όλες τις καταστάσεις που είναι προσβάσιμες με μηδέν ή περισσότερες ε-μεταβάσεις.

Ερώτημα

Πότε πρέπει το σύνολο-κατάσταση να θεωρείται τελικό;

Η ε-κλειστότητα μιας κατάστασης

Ορισμός

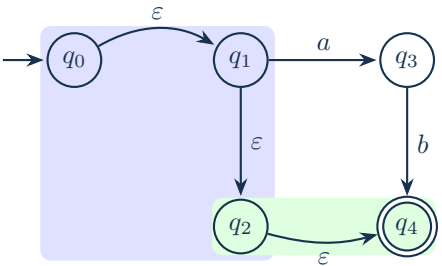
Για $q \in K$ ορίζουμε $E(q)$ ως το σύνολο όλων των καταστάσεων που φτάνουμε από το q χρησιμοποιώντας **μόνο** μεταβάσεις ε (ίσως καμία):

$$E(q) = \{p \in K : (q, \varepsilon) \vdash_M^* (p, \varepsilon)\}.$$

Για $Q \subseteq K$, θέτουμε $E(Q) = \bigcup_{q \in Q} E(q)$.

- Πάντα $q \in E(q)$: επιτρέπονται **μηδέν** μεταβάσεις.
- Το $E(q)$ είναι το **ελάχιστο** σύνολο που περιέχει το q και είναι κλειστό ως προς τις ε-ακμές.
- Υπολογισμός: ξεκινάμε από $\{q\}$ και προσθέτουμε επαναληπτικά τους ε-διαδόχους.

Παράδειγμα: ε-κλειστότητα σε ένα απλό ε-NFA



Ανάγνωση του σχήματος

Οι μπλε καταστάσεις δείχνουν το

$$E(q_0) = \{q_0, q_1, q_2, q_4\}.$$

Πράγματι:

$$q_0 \xrightarrow{\varepsilon} q_1 \xrightarrow{\varepsilon} q_2 \xrightarrow{\varepsilon} q_4.$$

Οι πράσινες καταστάσεις δείχνουν το

$$E(q_2) = \{q_2, q_4\}.$$

Εδώ από το q_2 μπορούμε:

- να μείνουμε στο q_2 (μηδέν κινήσεις),
- να πάμε με ε στο q_4.

Υπολογισμός της ε-κλειστότητας βήμα προς βήμα

Παράδειγμα για το $E(q_0)$

Ξεκινάμε από το σύνολο $\{q_0\}$.

$$\{q_0\} \Rightarrow \{q_0, q_1\} \Rightarrow \{q_0, q_1, q_2\} \Rightarrow \{q_0, q_1, q_2, q_4\}.$$

- Από το q_0 υπάρχει ε-μετάβαση στο q_1.
- Από το q_1 υπάρχει ε-μετάβαση στο q_2.
- Από το q_2 υπάρχει ε-μετάβαση στο q_4.
- Δεν προσθέτουμε το q_3, γιατί φτάνουμε εκεί μόνο με το σύμβολο a.

Άρα τελικά:

$$E(q_0) = \{q_0, q_1, q_2, q_4\}. \qquad E(q_2) = \{q_2, q_4\}.$$

Τυπική κατασκευή του ισοδύναμου DFA

Έστω $M = (K, \Sigma, \Delta, s, F)$. Ορίζουμε $D = (K', \Sigma, \delta', s', F')$ με

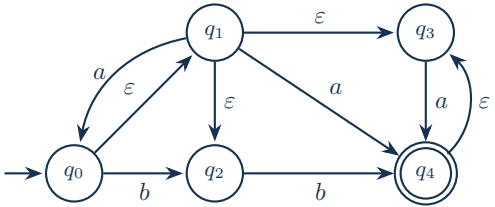
$$K' = \mathcal{P}(K),$$
$$s' = E(s),$$
$$F' = \{Q \subseteq K : Q \cap F \neq \varnothing\},$$
$$\delta'(Q, a) = E\left(\bigcup_{q \in Q} \{p \in K : (q, a, p) \in \Delta\}\right).$$

Με λόγια

$$\delta'(Q, a) = \underbrace{E}_{\varepsilon\text{-κλειστότητα}}\ \underbrace{\big(\mathrm{Move}(Q, a)\big)}_{\text{ένα γράμμα } a}.$$

- Η δ' επιστρέφει **ακριβώς ένα σύνολο** για κάθε (Q, a).
- Ακόμη και το $\varnothing$ είναι νόμιμη κατάσταση (παγίδα).

Παράδειγμα: το αρχικό ε-NFA (Παράδειγμα Lewis-Papadimitriou)



Πρώτα υπολογίζουμε τις κλειστότητες

$E(q_0) = \{q_0, q_1, q_2, q_3\}, \quad E(q_1) = \{q_1, q_2, q_3\},$
$E(q_2) = \{q_2\}, \quad E(q_3) = \{q_3\}, \quad E(q_4) = \{q_3, q_4\}.$

Παράδειγμα: οι πρώτες μεταβάσεις

Η αρχική κατάσταση του DFA είναι

$$A = E(q_0) = \{q_0, q_1, q_2, q_3\}.$$

Διαβάζουμε a

Από το A υπάρχουν a-ακμές προς q_0 και q_4. Άρα

$$\delta'(A, a) = E(\{q_0, q_4\}) = \{q_0, q_1, q_2, q_3, q_4\} =: B.$$

Διαβάζουμε b

Οι b-ακμές από το A καταλήγουν στα q_2 και q_4. Άρα

$$\delta'(A, b) = E(\{q_2, q_4\}) = \{q_2, q_3, q_4\} =: C.$$

Συγκρατούμε τα εξής:

Τα B, C είναι τελικές καταστάσεις, επειδή περιέχουν το q_4.

Παράδειγμα: συνεχίζουμε συστηματικά

Ονομάζουμε τα προσπελάσιμα υποσύνολα:

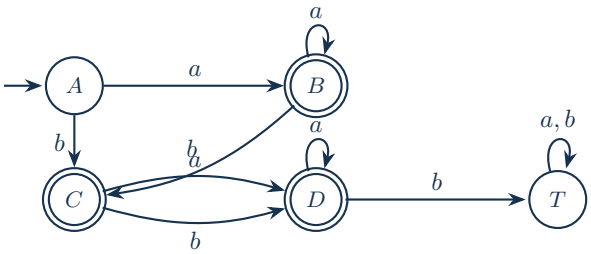
$$A = \{q_0, q_1, q_2, q_3\}, \quad B = \{q_0, q_1, q_2, q_3, q_4\},$$
$$C = \{q_2, q_3, q_4\}, \quad D = \{q_3, q_4\}, \quad T = \varnothing.$$

Κατάσταση	a	b	Τελική;
$\to A$	B	C	όχι
B	B	C	ναι
C	D	D	ναι
D	D	T	ναι
T	T	T	όχι

Παρατήρηση

Το πλήρες δυναμοσύνολο έχει $2^5 = 32$ στοιχεία, αλλά από την αρχική κατάσταση φτάνουμε μόνο σε **5** από αυτά.

Το DFA που προκύπτει



Τι αναπαριστά κάθε κόμβος;

$A = \{q_0, q_1, q_2, q_3\}$, $B = K$, $C = \{q_2, q_3, q_4\}$, $D = \{q_3, q_4\}$, $T = \varnothing$.

Παράδειγμα: υπολογισμός της $\delta'^*(A, aba)$

Ξεκινάμε από την αρχική κατάσταση του DFA:

$$A = E(q_0) = \{q_0, q_1, q_2, q_3\}.$$

Διαβάζουμε τη λέξη $w = aba$

Από τον πίνακα μεταβάσεων του DFA έχουμε:

$$\underbrace{A}_{\text{αρχή}} \xrightarrow{a} \underbrace{B}_{\delta'(A,a)} \xrightarrow{b} \underbrace{C}_{\delta'(B,b)} \xrightarrow{a} \underbrace{D}_{\delta'(C,a)}.$$

$$\delta'^*(A, \varepsilon) = A, \delta'^*(A, a) = \delta'(A, a) = B, \delta'^*(A, ab) = \delta'(B, b) = C,$$
$$\delta'^*(A, aba) = \delta'(C, a) = D.$$

Σύνδεση με το αρχικό ε-NFA

Επειδή $D = \{q_3, q_4\}$, καταλήγουμε στο $\boxed{\delta'^*(E(q_0), aba) = D = \{q_3, q_4\} = R(q_0, aba).}$

Απόδειξη ορθότητας: το βασικό αναλλοίωτο

Για κάθε $w \in \Sigma^*$ και κάθε κατάσταση $q \in K$, ορίζουμε

$$R(q, w) = \{p \in K : (q, w) \vdash_M^* (p, \varepsilon)\}.$$

Ισχυρισμός

Έστω δ'^* η επεκταμένη συνάρτηση μετάβασης του DFA.Αν ξεκινήσουμε το DFA από την $E(q)$ και διαβάσουμε τη λέξη w, καταλήγουμε **ακριβώς** στο σύνολο $R(q, w)$:

$$\boxed{\delta'^*(E(q), w) = R(q, w).}$$

- Η αριστερή πλευρά είναι **μία κατάσταση** του DFA, δηλαδή ένα υποσύνολο του K.
- Η δεξιά πλευρά είναι **όλες** οι πιθανές τελικές καταστάσεις του ε-NFA μετά την ανάγνωση του w.

Θα το αποδείξουμε με **επαγωγή στο μήκος** $|w|$.

Απόδειξη ισχυρισμού: βάση επαγωγής

Βάση: $w = \varepsilon$

Χωρίς ανάγνωση συμβόλου, το M μπορεί να ακολουθήσει μόνο μεταβάσεις ε. Επομένως

$$R(q, \varepsilon) = E(q).$$

Αντίστοιχα, το DFA δεν κάνει καμία μετάβαση όταν η είσοδος είναι κενή:

$$\delta'^*(E(q), \varepsilon) = E(q).$$

Άρα

$$\boxed{\delta'^*(E(q), \varepsilon) = R(q, \varepsilon).}$$

Συγκρατούμε τα εξής:

Η βάση εξηγεί γιατί η αρχική κατάσταση του DFA πρέπει να είναι $E(s)$ και όχι απλώς $\{s\}$.

Απόδειξη ισχυρισμού: επαγωγικό βήμα

Έστω ότι $\delta'^*(E(q), v) = R(q, v)$ για κάθε λέξη v μήκους k. Παίρνουμε $w = va$ με $a \in \Sigma$.

Διαχωρίζουμε την ανάγνωση της τελευταίας θέσης

Κάθε διαδρομή του M που διαβάζει va αποτελείται από:

1. μία διαδρομή που καταναλώνει v και τελειώνει σε κάποια $r \in R(q, v)$,
2. μία μετάβαση $(r, a, t) \in \Delta$,
3. μηδέν ή περισσότερες ε-μεταβάσεις από το t.

Άρα

$$R(q, va) = E\left(\bigcup_{r \in R(q,v)} \{t : (r, a, t) \in \Delta\} \right).$$

Χρησιμοποιώντας την επαγωγική υπόθεση και τον ορισμό της δ':

$$\begin{aligned} \delta'^*(E(q), va) &= \delta'(\delta'^*(E(q), v), a) \\ &= \delta'(R(q, v), a) = R(q, va). \end{aligned}$$

□

Ολοκλήρωση: τα δύο αυτόματα δέχονται την ίδια γλώσσα

Από τον ισχυρισμό, για κάθε λέξη w ισχύει

$$\delta'^*(s', w) = \delta'^*(E(s), w) = R(s, w).$$

$$\begin{aligned} w \in L(M) &\iff R(s,w) \cap F \neq \varnothing \\ &\iff \delta'^*(s', w) \in F' \\ &\iff w \in L(D). \end{aligned}$$

Συμπέρασμα

$$\boxed{L(M) = L(D).}$$

Κάθε ε-NFA μετατρέπεται αλγοριθμικά σε ισοδύναμο DFA.

Συγκρατούμε τα εξής:

Ο τύπος $F' = \{Q : Q \cap F \neq \varnothing\}$ αποτυπώνει το «αρκεί **μία** αποδεκτή διαδρομή».

Αλγόριθμος και πλήθος καταστάσεων

1. Υπολόγισε τις ε-κλειστότητες $E(q)$.
2. Ξεκίνα από $s' = E(s)$ και βάλε το στον κατάλογο ανεπεξέργαστων συνόλων.
3. Για κάθε νέο Q και κάθε $a \in \Sigma$, βρες $\delta'(Q, a) = E(\mathrm{Move}(Q, a))$.
4. Αν προκύψει νέο σύνολο, πρόσθεσέ το στον κατάλογο. Επανάλαβε ώσπου να μην υπάρχει νέο σύνολο.
5. Σήμανε ως τελικό κάθε Q με $Q \cap F \neq \varnothing$.

Πολυπλοκότητα ως προς το πλήθος καταστάσεων

Για $n = |K|$ υπάρχουν το πολύ 2^n διαφορετικά υποσύνολα. Επομένως η κατασκευή μπορεί να έχει **εκθετικό** πλήθος καταστάσεων· υπάρχουν οικογένειες ε-NFA για τις οποίες αυτό το εκθετικό μέγεθος είναι αναγκαίο.