

Automate fini non déterministe

Définition

Définition : (Automate fini non déterministe)

Un automate fini non déterministe (AFN) est un quintuplet $\mathcal{A} = (A, Q, I, E, T)$ où :

- A est un alphabet
- Q est un ensemble fini, l'ensemble des états
- $I \subseteq Q$ est l'ensemble des états initiaux
- $T \subseteq Q$ est l'ensemble des états finaux (ou terminaux)
- $E \subseteq Q \times (A \cup \{\varepsilon\}) \times Q$ est l'ensemble des relations de transition de \mathcal{A} (E définit une relation de transition sur \mathcal{A})

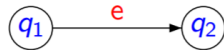
Remarques :

- Une transition est un triplet de E . Elle est de la forme (q_1, a, q_2) ou (q_1, ε, q_2) où q_1 et q_2 sont des états, a est une lettre de A et ε le mot vide.
- On note parfois une transition : $q_1 \xrightarrow{a} q_2$
- q_1 est l'origine de la transition, q_2 son extrémité et a et ε les étiquettes des transitions.

Représentation sagittale

Un AFN (A, Q, I, E, T) est représenté par un graphe orienté étiqueté tel que :

- Les sommets du graphe sont les états de l'automate ; ils sont cerclés
- Toute transition de la forme (q_1, e, q_2) est représentée par un arc étiqueté par $e \in A \cup \{\varepsilon\}$ et en reliant les sommets q_1 et q_2



- Les états initiaux sont marqués par une pointe de flèche (ou une flèche entrante)
- Les états terminaux sont doublement cerclés (ou marqués par des flèches sortantes non reliées à d'autres sommets)

Calcul dans un automate

Définition : (Calcul)

Soit $\mathcal{A} = (A, Q, I, E, T)$ un AFN, un calcul dans \mathcal{A} est une suite c de transitions $((q_i, a_i, q_{i+1}))_{i \in [1, n-1]}$ telle que l'extrémité d'une transition est l'origine de la suivante. On note un tel calcul de la façon suivante :

$$c = q_1 \xrightarrow{a_1} q_2 \xrightarrow{a_2} q_3 \xrightarrow{a_3} \dots \xrightarrow{a_{n-1}} q_n$$

On dit que :

- $q_1 \in Q$ est l'origine du calcul c
- $q_n \in Q$ est l'extrémité du calcul c
- $a_1 a_2 a_3 \dots a_{n-1} \in A^*$ est l'étiquette du calcul c .

On peut se permettre de noter $q \xrightarrow{\alpha} q'$ un calcul d'origine q , d'étiquette $\alpha \in A^*$ et d'extrémité q' sans expliciter les états intermédiaires.

Définition : (Calcul réussi)

On dit qu'un calcul dans $\mathcal{A} = (A, Q, I, E, T)$ est réussi lorsque son extrémité appartient à I et son extrémité appartient à T . Autrement dit :

$$c \text{ est un calcul réussi} \Leftrightarrow c = p \xrightarrow{\alpha} q, p \in I, q \in T, \alpha \in A^*$$

Définition : (État accessible)

Soit $\mathcal{A} = (A, Q, I, E, T)$ un AFN et $q \in Q$.

On dit que q est un état accessible si et seulement s'il existe un mot $\alpha \in A^*$ et un état initial $q_0 \in I$ tel que $q_0 \xrightarrow{\alpha} q$, autrement dit, s'il existe un calcul dont l'origine est un état initial et dont l'extrémité est q .

L'ensemble des états accessible de \mathcal{A} est donc :

$$\{q \mid q \in Q \text{ et } \exists \alpha \in A^*, \exists q_0 \in I, q_0 \xrightarrow{\alpha} q\}$$

Automate fini déterministe

Définition : (Automate fini déterministe))

Un automate fini déterministe (AFD) est un quintuplet $\mathcal{A} = (A, Q, q_0, \delta, T)$ tel que :

- A est un alphabet
- Q est un ensemble fini, l'ensemble des états
- $q_0 \in Q$ est l'unique état initial
- $\delta : Q \times A \rightarrow Q$ est une fonction appelée fonction de transition
- $T \subseteq Q$ est l'ensemble des états terminaux.

Remarques :

Soit $\mathcal{A} = (A, Q, I, E, T)$ un AFN, si \mathcal{A} comporte au moins l'une des caractéristiques suivantes :

- Plusieurs états initiaux
- Au moins une transition étiquetée par ε
- Un état qui est l'origine de plusieurs transitions différentes de même étiquette

alors on dit que l'automate \mathcal{A} est non déterministe, sinon on dit que \mathcal{A} est déterministe.

Définition : (Prolongement de δ)

Soit (A, Q, δ, s_0, F) un AFD, la fonction de transition $\delta : Q \times A \rightarrow Q$ se prolonge en une application $\delta^* : Q \times A^* \rightarrow Q$ définie par :

- $\delta^*(q, \varepsilon) = q$ pour tout q de Q
- $\delta^*(q, \alpha.a) = \delta(\delta^*(q, \alpha), a)$ pour tout q de Q , pour tout α de A^* et tout a de A

Remarques et propriété :

- δ^* est définie par récurrence sur la longueur du mot
- δ^* est un prolongement de δ (c'est-à-dire $\delta^*(q, a) = \delta(q, a)$ pour tout état q et toute lettre a)
- Ce prolongement vérifie $\delta^*(q, \alpha.\beta) = \delta^*(\delta^*(q, \alpha), \beta)$ pour tout état q et tout mots α et β (On le prouverait par récurrence sur la longueur de β)

Langage reconnu par un automate fini et langage reconnaissable

Définition : (Mot reconnu)

Soit \mathcal{A} un automate fini, défini sur un alphabet A .

Un mot α de A^* est reconnu (ou accepté) par \mathcal{A} s'il existe un calcul réussi d'étiquette α

Définition : (Langage reconnu par un automate fini)

Soit $\mathcal{A} = (A, Q, I, E, T)$.

Le langage reconnu par \mathcal{A} , que l'on note $\mathcal{L}(\mathcal{A})$ est l'ensemble des mots reconnus par \mathcal{A} :

$$\mathcal{L}(\mathcal{A}) = \{\alpha \mid \alpha \in A^* \text{ et } \exists q \in I, \exists q' \in T, q \xrightarrow{\alpha} q'\}$$

Définition : (Langage reconnaissable)

On dit qu'un langage est reconnaissable par un automate fini s'il existe un automate fini qui le reconnaît. On note $Rec(A^*)$ l'ensemble des langages sur l'alphabet A reconnaissables par un automate fini

Remarques :

- Étant donné un mot et un AFD il est facile de vérifier si le mot est reconnu en utilisant δ (Ou de façon pratique avec la représentation sagittale)
- Soit $\mathcal{A} = (A, Q, I, E, T)$ un AFD :
 - Un mot α de A^* est reconnu par \mathcal{A} si et seulement si $\delta^*(q_0, \alpha) \in T$
 - $\mathcal{L}(\mathcal{A}) = \{\alpha \mid \alpha \in A^* \text{ et } \exists q' \in T, \delta^*(q_0, \alpha) = q'\}$
- Si l'automate est non déterministe il faut simuler toutes les exécutions possibles...

Théorème de Kleene :

L'ensemble $Rat(A^*)$ des langages réguliers (ou rationnels) est égal à l'ensemble $Rec(A^*)$ des langages reconnus par un automate fini :

$$Rat(A^*) = Rec(A^*)$$

La démonstration se fait par double inclusion (cf. diaporama)

Langage associé à un état

Définition :

Soit $\mathcal{A} = (A, Q, I, E, T)$ un automate fini et $p, q \in Q$

- On note $E_{p,q}$ l'ensemble des étiquettes des transitions d'origine p et d'extrémité q
- On pose $L_p = \{\alpha \mid \alpha \in A^* \text{ et } \exists r \in T, p \xrightarrow{\alpha} r\}$

Remarques :

- L_p est l'ensemble des mots qui sont les étiquettes des calculs d'origine p et d'extrémité un état final de \mathcal{A}
- A chaque état $p \in Q$ on peut associer un langage L_p
- On a évidemment $\mathcal{L}(\mathcal{A}) = \sum_{p \in I} L_p$

Système d'équations

Pour montrer que $\mathcal{L}(\mathcal{A})$ est rationnel il suffit de montrer que chacun des L_p est rationnel.

Si $\alpha \in L_p$ on a 2 cas :

- Soit $\alpha = \varepsilon$ dans ce cas $p \in T$
- Soit $\alpha = a.\beta$ avec a l'étiquette d'une transition de \mathcal{A} qui va de p à q et $\beta \in L_q$

On peut donc écrire :

$$\forall p \in Q, L_p = \sum_{q \in Q} E_{p,q} L_q + \delta_{p,T} \quad (E_1)$$

où $\delta_{p,T} = \{\varepsilon\}$ si $p \in T$ et \emptyset sinon. (E_1) définit donc un système de n équations linéaires à n inconnues avec $n = \text{card}(Q)$, les inconnues étant les L_p , et les coefficients dans $\mathcal{P}(A)$.

La résolution d'un tel système nécessite parfois l'utilisation du lemme d'Arden.

Lemme d'Arden :

Soient A et B deux langages définis sur l'alphabet V , on considère l'équation $X = A.X + B$ (E_2) où l'inconnue X est un langage sur V

- Si $\varepsilon \notin A$, l'équation (E_2) admet une solution unique $A^*.B$
- Si $\varepsilon \in A$, les solutions de l'équation (E_2) sont $A^*.L$ où $B \subset L \subset V^*$

Théorème :

Soit le système suivant de $n > 0$ équations :

$$\left\{ L_i = \sum_{j=0}^n E_{i,j}.L_j + F_i \right\}_{i \in \llbracket 0, n \rrbracket}$$

tels que les langages $E_{i,j} (1 \leq i, j \leq n)$ ne contiennent pas ε et les langages F_i sont les inconnues. Ce système admet une unique solution (et si les langages $E_{i,j}$ et F_i sont réguliers alors les L_i le sont aussi).