Langages d'images

Devoir à rendre au plus tard le 11 mai

Le but de ce devoir est l'étude de deux formalismes permettant de définir des langages d'images, c'est-à-dire des ensemble de matrices à deux dimensions sur un alphabet fini fixé. La première partie contient la définition des images, et des opérations de base pour les manipuler. La deuxième partie établit une classe de langages d'images définis par un système de pavage, tandis que la troisième partie introduit un modèle d'automates opérant sur des matrices. Les deux dernières parties sont largement indépendantes.

Préliminaires

Définition 1 (Image)**.** Une image sur un alphabet fini Σ est une matrice finie sur Σ, $m = (m_{i,j})_{0 \leq i < h, 0 \leq j < \ell}$. On appelle h la hauteur de m, et ℓ sa largeur. On note $h \times \ell$ la dimension de m. Finalement, on note Σ^{**} l'ensemble des images sur Σ.

Définition 2 (Concaténation)**.** Soit $m = (m_{i,j})_{0 \leq i < h, 0 \leq j < \ell}$ et $m' = (m'_{i,j})_{0 \leq i < h', 0 \leq j < \ell}$ deux images de même largeur ℓ. On note $m \boxminus m'$ la concaténation verticale de m et m', définie comme la matrice de dimension $(h + h') \times \ell$ suivante :

$$(m \boxminus m')_{i,j} = \begin{cases} m_{i,j} & \text{si } i < h \\ m'_{i-h,j} & \text{si } i \geq h \end{cases}$$

On définit de façon similaire la concaténation horizontale $m \boxbar m'$ de deux images de dimensions respectives $h \times \ell$ et $h \times \ell'$, ayant la même hauteur h comme :

$$(m \boxbar m')_{i,j} = \begin{cases} m_{i,j} & \text{si } j < \ell \\ m'_{i,j-\ell} & \text{si } j \geq \ell \end{cases}$$

Définition 3 (Image délimitée)**.** Soit m une image de dimension $h \times \ell$ sur Σ. Soit $\top, \perp, \vdash$ et $\dashv$ quatre symboles n'apparaissant pas dans Σ. On appelle image délimitée correspondant à m l'image $\boxed{m}$ de dimension $(h+2) \times (\ell+2)$ sur $\Sigma \cup \{\top, \perp, \vdash, \dashv\}$ définie par :

$$\boxed{m}_{i,j} = \begin{cases} \perp & \text{si } i = 0 \\ \top & \text{si } i = h+1 \\ \dashv & \text{si } j = 0 \text{ et } 0 < i < h+1 \\ \vdash & \text{si } j = \ell+1 \text{ et } 0 < i < h+1 \\ m_{i-1,j-1} & \text{sinon} \end{cases}$$

Exemple 1. On considère les images suivantes :

$$m_1 = \begin{pmatrix} a & b \\ a & a \end{pmatrix} \quad m_2 = \begin{pmatrix} b & a \\ b & a \\ a & b \end{pmatrix} \quad m_3 = \begin{pmatrix} b & a & a \\ a & b & a \end{pmatrix}$$

On a alors les égalités suivantes :

$$m_1 \boxminus m_2 = \begin{pmatrix} a & b \\ a & a \\ b & a \\ b & a \\ a & b \end{pmatrix} \quad m_1 \boxplus m_3 = \begin{pmatrix} a & b & b & a & a \\ a & a & a & b & a \end{pmatrix} \quad \boxed{m_1} = \begin{pmatrix} \perp & \perp & \perp & \perp \\ \dashv & a & b & \vdash \\ \dashv & a & a & \vdash \\ \top & \top & \top & \top \end{pmatrix}$$

Et on peut remarquer par exemple que la concaténation horizontale de m_1 avec m_2 n'est pas définie.

1 Reconnaissance par pavages

Soit Γ un alphabet fini. Soit $m \in \Gamma^{**}$. On note $\mathcal{P}_{p\times q}(m)$ l'ensemble des sous-images de m de dimension $p \times q$. Soit $\mathcal{P}$ un ensemble d'images de dimension $p \times q$ sur Γ. On dit que $\mathcal{P}$ engendre m si et seulement si $\mathcal{P}_{p\times q}(m) \subseteq \mathcal{P}$.

Un système de pavage de dimension $p\times q$ sur l'alphabet Σ est un 3-uplet $\mathcal{S} = \langle \Gamma, \pi, \mathcal{P}\rangle$, où Γ est un alphabet fini, π est un renommage qui envoie chaque lettre de Γ sur une lettre de Σ et $\mathcal{P}$ est un ensemble d'images de dimension $p \times q$ sur $\Gamma \cup \{\top, \perp, \vdash, \dashv\}$. On définit alors $\mathcal{L}(\mathcal{S})$, le langage reconnu par $\mathcal{S}$, comme :

$$\mathcal{L}(\mathcal{S}) = \{\pi(m) \mid m \in \Gamma^{**} \text{ et } \mathcal{P}_{p\times q}(\boxed{m}) \subseteq \mathcal{P}\}$$

Un langage d'images L est reconnu par pavage s'il existe un système de pavage $\mathcal{S}$ tel que $L = \mathcal{L}(\mathcal{S})$.

1. Soit $\Sigma = \{a, b\}$. Construire des systèmes de pavage de dimension 2×2 sur Σ reconnaissant les langages suivants :
   – l'ensemble des images dont la première et la dernière lignes sont des a ;
   – l'ensemble des images carrées (ie. dont la largeur est égale à la hauteur) ;
   – l'ensemble des images carrées de hauteur impaire dont la ligne du milieu est constituée uniquement de b.
2. Soit $\mathcal{S}$ un système de pavage sur Σ de dimension $p \times q$. Montrer qu'il existe un système de pavage équivalent $\mathcal{S}'$ de dimension 2×2.
3. Soit $\mathcal{S}$ un système de pavage sur Σ tel que toutes les images de $L = \mathcal{L}(\mathcal{S})$ sont de hauteur 1. Montrer alors que L est en fait un langage régulier de mots. Réciproquement, soit L un langage régulier, montrer que L, vu comme un ensemble d'images de hauteur 1, est reconnu par pavage.
4. Montrer que l'ensemble des langages reconnus par pavage est clos par union, intersection, concaténation verticale et concaténation horizontale.

On considère le langage suivant :

$$L = \{m \in \Sigma^{**} \mid \exists s \in \Sigma^{**}, m = (s \boxminus s) \text{ et } s \text{ est une image carrée.}\}$$

5. Montrer que si $\Sigma = \{a\}$, alors L est reconnu par pavage.
6. Montrer que si $\Sigma = \{a, b\}$, alors L n'est pas reconnaissable par pavage.
7. Montrer que le complémentaire de L est reconnu par pavage.

2 Automates d'images

Soit Σ un alphabet fini. Un automate d'images sur Σ est un tuple $\mathcal{A} = \langle Q, \delta, I, F\rangle$, où Q est un ensemble fini d'états, $I \subseteq Q$ est l'ensemble des états initiaux, $F \subseteq Q$ est l'ensemble des états finals, et δ est l'ensemble des transitions de l'automate, défini par une relation de $Q \times (\Sigma \cup \{\top, \bot, \vdash, \dashv\})$ dans $Q \times \{(-1,0), (+1,0), (0,0), (0,-1), (0,+1)\}$.

De plus, δ vérifie les propriétés suivantes, pour tout état $q \in Q$:

$$\delta(q, \bot) \subseteq Q \times \{(+1,0)\} \qquad \delta(q, \top) \subseteq Q \times \{(-1,0)\}$$

$$\delta(q, \dashv) \subseteq Q \times \{(0,+1)\} \qquad \delta(q, \vdash) \subseteq Q \times \{(0,-1)\}$$

Intuitivement, à chaque étape de calcul, un tel automate lit la lettre de l'image à la position où il se trouve, et, comme décrit par sa table de transition, change d'état et effectue un déplacement vers le bas $(+1, 0)$, vers le haut $(-1, 0)$, vers la droite $(0, +1)$, vers la gauche $(0, -1)$, ou bien reste sur place $(0, 0)$. Un calcul est acceptant s'il commence dans un état initial et à la position $(0, 0)$ de l'image, et atteint un état final. Les contraintes sur δ assurent que la tête de lecture de l'automate ne peut pas sortir de l'image.

Formellement, une configuration de $\mathcal{A}$ sur une image m de dimension $h \times \ell$ est un couple (q, p) ou $q \in Q$ est un état de l'automate, et $p \in \{0, \ldots, h-1\} \times \{0, \ldots, \ell-1\}$ est une position sur l'image. Une exécution de $\mathcal{A}$ sur m est une suite de configurations $\rho = (q_0, p_0) \ldots (q_k, p_k)$ telle que $q_0 \in I$, $p_0 = (0, 0)$ et pour tout $i < k$:

$$\exists (q, d) \in \delta(q_i, m_{p_i}), q_{i+1} = q \text{ et } p_{i+1} = p_i + d$$

Une telle exécution est acceptante si $q_k \in F$. Soit $m \in \Sigma^{**}$, on dit que m est acceptée par $\mathcal{A}$ si $\mathcal{A}$ a une exécution acceptante sur $\boxed{m}$. On note $\mathcal{L}(\mathcal{A})$ l'ensemble des images acceptées par $\mathcal{A}$. Finalement, on dit que $\mathcal{A}$ est déterministe si $|I| \leq 1$ et que δ est une fonction.

1. Construire des automates d'images reconnaissant les langages suivants :
   - l'ensemble des images carrées ;
   - l'ensemble des images dont la première ligne est égale à la dernière ligne ;
   - l'ensemble des images contenant la lettre c.
2. Soit $\mathcal{A}$ un automate d'images. Montrer qu'il existe un automate d'images $\mathcal{A}' = \langle Q', \delta', I', \{q'_f\}\rangle$ tel que :
   - $\mathcal{L}(\mathcal{A}) = \mathcal{L}(\mathcal{A}')$;
   - $\forall \alpha \in (\Sigma \cup \{\top, \bot, \vdash, \dashv\})$, $\delta'(q'_f, \alpha) = \emptyset$;
   - toutes les exécutions acceptantes de $\mathcal{A}'$ terminent à la position $(1, 1)$.

Montrer de plus que si $\mathcal{A}$ est déterministe, alors $\mathcal{A}'$ peut être choisi déterministe.

Dans toute la suite, on supposera que les automates d'images ont la propriété de la question 2.

3. Montrer que l'ensemble des langages reconnus par des automates d'images est clos par union et intersection.
4. Soit $\mathcal{A}$ un automate d'images déterministe et $m \in \Sigma^{**}$. On considère le graphe orienté $G_{\mathcal{A},m}$ dont les noeuds sont les configurations de $\mathcal{A}$ sur $\boxed{m}$, et tel que deux noeuds x et y sont connectés par $x \to y$ si et seulement si y est le successeur de x dans le calcul de $\mathcal{A}$ sur $\boxed{m}$. Soit G la composante connexe de $G_{\mathcal{A},m}$ contenant $(q_f, (1,1))$. Montrer que :
   - G est fini ;
   - G ne contient pas de boucles ;
   - G est en fait un arbre enraciné en $(q_f, (1,1))$.
5. En utilisant la question 4, montrer que l'ensemble des langages reconnus par des automates d'images déterministes est clos par complémentaire.
6. Montrer que l'ensemble des langages reconnus par des automates d'images déterministes est clos par union et par intersection.

Jusqu'à la fin du problème, on suppose que $\Sigma = \{a\}$. On pose :

$$L = \{m \in \Sigma^{**} \mid m \text{ est de dimension } h \times \ell \text{ et } \exists s, t \in \mathbb{N}, 0 \leq s \leq t \text{ et } \ell = th + s\}$$

et $\overline{L} = \Sigma^{**} - L$.

7. Montrer que L est reconnaissable par un automate d'images.
8. Soit $h \geq 3$, montrer que :
   - $\overline{L}$ contient un nombre fini d'images de hauteur h ;
   - $\overline{L}$ contient une image de dimension $h \times (h(h-1)-1)$.
9. En déduire que $\overline{L}$ n'est pas reconnaissable par un automate d'images.
10. Conclure que L n'est pas reconnaissable par un automate d'images déterministe, et que l'ensemble des langages reconnus par des automates d'images n'est pas clos par complémentaire.