Transductions rationnelles

Daphné Dieuleveut

19 décembre 2008

On s'intéresse dans le cours sur les automates aux propriétés des langages rationnels sur un alphabet fini, donc parties d'un monoïde libre A^*. Or, il est intéressant de travailler sur ces notions dans le cadre plus général de monoïdes finiment engendrés, voire quelconques. Le cas particulier d'un monoïde produit $A^* \times B^*$, dont l'étude a pu être motivée par l'analyse syntaxique et l'idée de créer des traducteurs automatiques, offre en lui-même une théorie relativement riche, fondée sur les transductions rationnelles.
Dans ce dossier, on se concentrera dans un premier temps sur les relations entre ensembles rationnels et reconnus par un monoïde fini, puis sur les propriétés des transductions rationnelles, et en particulier la reprsentation équivalente par des transducteurs.

1 Ensembles rationnels et reconnaissables dans un monoïde

1.1 Définitions et premières propriétés

On va dans un premier temps élargir les définitions des langages réguliers et rationnels sur un alphabet A^* à des parties d'un monoïde quelconque.
On rappelle tout d'abord quelques définitions relatives aux monoïdes :

Définition 1. Un monoïde est un ensemble $(M, \cdot)$ tel que $\cdot$ est une loi de composition associative, admettant un élément neutre 1_M, sur l'ensemble M.
M est dit finiment engendré s'il est engendré par un nombre fini d'éléments.
Un morphisme de monoïdes est une application $\mu : M \to N$, avec M, N monoïdes, telle que :
(1) $\mu(1_M) = 1_N$
(2) $\forall x, y \in M \ \mu(x \cdot y) = \mu(x) \cdot \mu(y)$.
M est dit libre s'il est isomorphe à A^*, avec A un alphabet (i.e. un ensemble fini de lettres).

On donne maintenant une définition d'une partie reconnaissable pour un monoïde quelconque.

Définition 2. Partie reconnaissable d'un monoïde.
Pour tout monoïde M, un ensemble $X \subset M$ est dit reconnaissable s'il existe un monoïde fini N, un morphisme de monoïdes $\mu : M \to N$ et une partie P de N tels que $X = \mu^{-1}(P)$.
On note $Rec(M)$ l'ensemble des parties reconnaissables de M.

Voyons une première propriété de clôture de $Rec(M)$.

Proposition 3. *Pour tout monoïde M, $Rec(M)$ est clos par union, intersection et complémentation.*

Démonstration. Soient $X_1,\ X_2 \in Rec(M)$.
Par la définition, on a existence de morphismes de monoïdes $\mu_1 : M \to N_1$, $\mu_2 : M \to N_2$, avec N_1, N_2 monoïdes finis, et de $P_1 \subset N_1$, $P_2 \subset N_2$ tels que $X_1 = \mu^{-1}(P_1)$, $X_2 = \mu^{-1}(P_2)$.
On définit un morphisme $\mu = (\mu_1, \mu_2) : M \to N = N_1 \times N_2$.
N est un monoïde fini, et $X_1 \cap X_2 = \mu^{-1}(P_1 \times P_2)$, donc $X_1 \cap X_2 \in Rec(M)$.
De plus, $\mu_1^{-1}(N_1 \setminus P_1) = M \setminus X_1$, donc $M \setminus X_1 \in Rec(M)$.
Comme $X_1 \cup X_2 = M \setminus ((M \setminus X_1) \cup (M \setminus X_2))$, on en déduit $X_1 \cup X_2 \in Rec(M)$. $\square$

Passons aux parties rationnelles :

Définition 4. Partie rationnelle d'un monoïde.
Pour tout monoïde M, on note $Rat(M)$ le plus petit sous-ensemble R de $\mathcal{P}(M)$ tel que :
(1) $\emptyset \in R$ et $\forall m \in M, \{m\} \in R$
(2) $\forall X, Y \in R : X \cup Y,\ X \cdot Y,\ X^* \in R$. (propriété de clôture pour les opérations rationnelles)
$Rat(M)$ est l'ensemble des parties rationnelles de M.

Le théorème de Kleene assure que, pour un alphabet A, $Rec(A^*) = Rat(A^*)$. Ce résultat n'est pas vrai dans des monoïdes quelconques ; on verra un contre-exemple dans la preuve de la proposition 13. Lorsque $X \in Rat(M) \cap Rec(M)$, on dit que X est régulier.

On va maintenant s'intéresser à un résultat comparable au théorème de Kleene, dans le cadre des monoïdes finiment engendrés.

1.2 Une relation entre $Rec(M)$ et $Rat(M)$

On va dans un premier temps montrer quelques résultats utiles de stabilité du caractère reconnaissable, puis rationnel, d'un ensemble, par morphisme ou morphisme inverse.

Proposition 5. *(Clôture par morphisme inverse.) Soient M, M' monoïdes, $\mu : M \to M'$ morphisme. Alors $\forall X' \subset M'$*

$$X' \in Rec(M') \Rightarrow \mu^{-1}(X') \in Rec(M).$$

Démonstration. Soit $X' \in Rec(M')$.
Il existe N monoïde fini, $\nu : M \to N$ morphisme de monoïdes et $P \subset N$ tels que $X = \nu^{-1}(P)$.
Alors $\mu^{-1}(X') = (\nu \circ \mu)^{-1}(P) \in Rec(M)$. $\square$

Ce résultat n'assure évidemment pas que pour un morphisme $\mu : M \to M'$, l'implication $\mu(X) \in Rec(M') \Rightarrow X \in Rec(M)$ soit vérifiée (a priori $\mu^{-1}(\mu(X)) \neq X$).
On en déduit toutefois une propriété de clôture de $Rec(M)$ par isomorphisme :

Corollaire 6. *Soient M, M' monoïdes, $\mu : M \to M'$ isomorphisme. Alors $\forall X \subset M$*

$$X \in Rec(M) \Leftrightarrow \mu(X) \in Rec(M').$$

Démonstration. En notant ν le morphisme inverse de μ, on a $\nu^{-1}(X) = \mu(X)$ (au sens de l'image réciproque) ; et $\mu^{-1}(\mu(X)) = X$. $\square$

On va montrer que $Rat(M)$ vérifie une propriété de clôture similaire, à l'aide de la proposition suivante :

Proposition 7. *Soient M, M' monoïdes, $\mu : M \to M'$ morphisme.*
Alors $\forall X \in Rat(M),\ \mu(X) \in Rat(M')$.

Démonstration. Il suffit de montrer que $\mathcal{A} = \{X \in Rat(M)/\ \mu(X) \in Rat(M')\}$ est clos pour les opérations rationnelles et contient $\emptyset$ et les singletons $\{m\}, m \in M$(la proposition résulte alors de la définition de $Rat(M)$).
On a bien $\emptyset \in Rat(M)$, et $\forall m \in M, \mu(\{m\}) = \{\mu(m)\} \in Rat(M')$ donc $\{m\} \in \mathcal{A}$. Et pour $X,\ Y \in \mathcal{A}$: $\mu(X \cup Y) = \mu(X) \cup \mu(Y),\ \mu(XY) = \mu(X)\mu(Y)$ et $\mu(X^+) = (\mu(X))^+$ donc $X \cup Y, XY, X^+ \in \mathcal{A}$. $\square$

Corollaire 8. *Si $\mu : M \to M'$ isomorphisme de monoïdes, alors $\forall X \subset M$*

$$X \in Rat(M) \Leftrightarrow \mu(X) \in Rat(M').$$

Appliquons ces résultats au cas d'un monoïde finiment engendré, pour montrer une inclusion entre $Rec(M)$ et $Rat(M)$.

Théorème 9. *Soit M monoïde finiment engendré. Alors $Rec(M) \subset Rat(M)$.*

Démonstration. M est finiment engendré, donc il existe un morphisme surjectif $\mu : A^* \to M$, avec A alphabet (il suffit de prendre $A = \{1, ..., n\},\ \mu(k) = m_k$, avec $\{m_k,\ k = 1, ..., n\}$).
Soit $X \in Rec(M)$. On a $\mu^{-1}(X) \in Rec(A^*)$ d'après la proposition 5, donc par le théorème de Kleene $\mu^{-1}(X) \in Rat(A^*)$, d'où par la proposition 7 :
$\mu(\mu^{-1}(X)) \in Rat(M)$.
Comme μ est surjectif, $\mu(\mu^{-1}(X)) = X$. $\square$

On va à présent s'intéresser aux propriétés des parties reconnaissables et rationnelles d'un monoïde produit. Pour cela, on fait appel à la notion de transduction.

2 Transductions rationnelles

2.1 Relations reconnaissables et rationnelles, transductions

Définition 10. Soient M, N monoïdes.
Une transduction est un application $\tau : M \to \mathcal{P}(N)$.
Le domaine et l'image de τ sont définis par :
$dom(\tau) = \{x \in M / \tau(x) \neq \emptyset\}$
$im(\tau) = \{y \in N / \exists x \in M, y \in \tau(x)\}$.
La relation binaire associée à τ est :
$\mathcal{R} = \{(x, y) \in M \times N / y \in \tau(x)\}$.

On notera pour simplifier $\tau : M \to N$ au lieu de $\tau : M \to \mathcal{P}(N)$.

Définition 11. Relation, transduction rationnelles.
Une relation rationnelle est une partie rationnelle de $M \times N$, avec M, N monoïdes. (On definit de même une relation reconnaissable.)
Une transduction $\tau : M \to N$ est dite rationnelle si la relation associée est rationnelle.

Pour des transductions $\tau_1, \tau_2 : M \to N$, on définit $\tau_1 \cup \tau_2$, $\tau_1\tau_2$ et τ_1^+ comme suit : $\forall x \in M$

$$\tau_1 \cup \tau_2(x) = \tau_1(x) \cup \tau_2(x)$$
$$\tau_1\tau_2(x) = \tau_1(x)\tau_2(x)$$
$$\tau_1^+(x) = \bigcup_{n=0}^{\infty} \tau_1(x)^n$$

Proposition 12. *Si τ_1, τ_2 sont des transductions rationnelles, alors les transductions $\tau_1 \cup \tau_2$, $\tau_1\tau_2$ et τ_1^+ sont aussi rationnelles.*

Démonstration. Les relations associées à $\tau_1 \cup \tau_2$, $\tau_1\tau_2$ et τ_1^+ sont (respectivement) $\mathcal{R}_1 \cup \mathcal{R}_2$, $\mathcal{R}_1\mathcal{R}_2$ et $\mathcal{R}_1^+$ (avec $\mathcal{R}_1$, $\mathcal{R}_2$ relations associées à τ_1, τ_2), donc sont rationnelles. $\square$

On s'intéresse en particulier au cas de transductions $\tau : A^* \to B^*$, avec A, B alphabets.

Proposition 13. $Rec(A^* \times B^*) \subsetneq Rat(A^* \times B^*)$

Démonstration. L'inclusion résulte du théorème 9. Voyons un contre-exemple à l'inclusion réciproque.
On considère $X = \{(a^n, b^n) / n \in \mathbb{N}\}$, avec $a \in A$, $b \in B$.
$X = \{(a, b)\}^* \in Rat(A^* \times B^*)$
Soit C l'alphabet $\{a, b\}$, $\mu : C^* \to A^* \times B^*$ le morphisme défini par :
$\mu(a) = (a, 1)$, $\mu(b) = (1, b)$ (avec 1 le mot vide).
$\mu^{-1}(X) = \{u \in C^* / \mid u \mid_a = \mid u \mid_b\} \notin Rec(C^*)$ donc (contraposée de la proposition 5) $X \notin Rec(A^* \times B^*)$. $\square$

On va voir que les relations rationnelles (et donc les transductions rationnelles) sont en fait représentés par une généralisation des automates, les transducteurs.

2.2 Transducteurs

Définition 14. Transducteur.
Un transducteur $\mathcal{T} = \langle A, B, Q, q_0, F, E\rangle$ se compose de :
- deux alphabets A, B
- un ensemble fini d'états Q
- un état initial $q_0 \in Q$
- un ensemble d'états finaux $F \subset Q$
- un ensemble de transitions $E \subset Q \times A^* \times B^* \times Q$

A est appelé alphabet d'entrée, B alphabet de sortie.

Dans les définitions suivantes, on se fixe un transducteur $\mathcal{T} = \langle A, B, Q, q_0, F, E\rangle$.

Définition 15. Calcul, calcul acceptant.
Un calcul dans un tel transducteur est une suite d'états $c_0 \xrightarrow{(a_1,b_1)} c_1 \cdots \xrightarrow{(a_n,b_n)} c_n$ telle que pour $i = 1, ..., n$, $(c_{i-1}, a_i, b_i, c_i) \in E$. $(a_1...a_n,\ b_1...b_n)$ est l'étiquette du calcul.
Le calcul est dit acceptant si $c_0 = q_0$ et $c_n \in F$.

Pour tous états $p, q \in Q$, on note $\Lambda(p, q)$ l'ensemble des étiquettes de calculs de p à q dans $\mathcal{T}$; et pour tout $Q' \subset Q$, $\Lambda(p, Q') = \{\Lambda(p, q)/q \in Q\}$.

Définition 16. Transduction associée à $\mathcal{T}$.
La transduction $\tau : A^* \to B^*$ associée à $\mathcal{T}$ est définie par :
$\tau(x) = \{y \in B^*/(x, y) \in \Lambda(q_0, F)\}$.

Ces définitions permettent d'établir une caractérisation des transductions rationnelles par les transducteurs. Pour montrer cette propriété, on cherche à se ramener aux résultats connus sur les langages rationnels et la reconnaissance par automates finis. On s'aidera de la proposition suivante (admise dans ce dossier ; la démonstration résulte de caractérisations techniques du caractère rationnel d'une relation à l'aide de morphismes) :

Proposition 17. *Soient M, M' monoïdes. Une relation X sur $M \times M'$ est rationnelle si et seulement s'il existe un alphabet C, deux morphismes $\mu : C^* \to M$, $\mu' : C^* \to M'$ et un langage régulier K sur C tels que $X = \{(\mu(w), \mu'(w))/w \in K\}$.*

On a alors de manière immédiate le résultat suivant :

Corollaire 18. *Soient M, M' monoïdes. Une transduction $\tau : M \to M'$ est rationnelle si et seulement s'il existe un alphabet C, deux morphismes $\mu : C^* \to M$, $\mu' : C^* \to M'$ et un langage régulier K sur C tels que $\forall x \in M$, $\tau(x) = \mu'(\mu^{-1}(x) \cap K))$*

On peut alors énoncer le théorème suivant :

Théorème 19. *Une transduction est rationnelle si et seulement si c'est la transduction associée à un transducteur.*

Démonstration. - Transducteur $\rightsquigarrow$ transduction rationnelle :
Soit $\mathcal{T} = \langle A, B, Q, q_0, F, E\rangle$ un transducteur.
$\forall p \in Q,\ Q \subset Q'$, on a $Lambda(p, Q') \in Rec(A^* \times B^*)$:
En effet $Lambda(p, q) = ((\Lambda(p, q) \cap 1) \cup (U_p E^* \cap E^* V_p)) \setminus E^* W E^*$
avec $U_p = \{(p, u, v, p') \in E\},\ V_q = \{(q', u, v, q) \in E\}$,
$W = \{(q_1, u_1, v_1, q_1')(q_2, u_2, v_2, q_2') \in E^2 / q_1' \neq q_2\}$
Donc $\Lambda(p, Q')$ est reconnaissable.
On considère les morphismes $\mu : E^* \to A^*,\ \nu : E^* \to B^*$, qui à un calcul associent les projections canoniques de son étiquette sur A^*, B^*. On a alors :
$\forall x \in A^*,\ \tau(x) = \nu(\mu^{-1}(x))$ Donc τ est un transduction rationnelle (d'après la proposition).
- Transduction rationnelle $\rightsquigarrow$ transducteur :
Soit $\tau : A^* \to B^*$ une transduction rationnelle. On peut supposer A et B disjoints (quitte à envoyer bijectivement B sur un alphabet B' disjoint de A)
On a donc $\forall X \in A^*,\ \tau(x) = \pi_B(\pi_A^{-1} \cap K)$
avec K langage régulier sur $(A \cup B)^*$, et $\pi_A,\ \pi_B$ projections de $(A \cup B)^*$ sur $A^*,\ B^*$.
K est reconnu par un automate $\mathcal{A} = \langle A \cup B, Q, q_0, F\rangle$.
On considère le transducteur $\mathcal{T} = \langle A, B, Q, q_0, F, E\rangle$,
avec $E = \{(q, \pi_A(w), \pi_B(w), q \cdot w)/q \in Q, w \in A \cup B\}$.
Alors τ est la transduction associée à $\mathcal{T}$. $\square$

Bibiographie : Jean Berstel, *Transductions and context-free languages*