Le Problème de Séparation par Automate est NP-Complet

24 octobre 2011

Théorème 1

Instance : $k \in \mathbb{N}, S, T \subset \Sigma^*$ *des langages finis*

Problème PSA : $\exists ? \mathcal{A}$ *un AFD à k états tels que* $S \subset \mathcal{L}(\mathcal{A})$ *et* $T \cap \mathcal{L}(A) = \varnothing$

Théorème : *PSA est un problème NP-complet.*

Démonstation :

- *PSA est NP :* On sais créer en temps polynomial un automate à k état et vérifier s'il sépare S et T.
- *PSA est NP-difficile :* On va réduire le problème SAT au problème PSA. Comme SAT est NP-Complet, par le *théorème de Cook*, PSA sera aussi NP-Complet.

Construisons donc la réduction de SAT à PSA :
Soit F une formule :

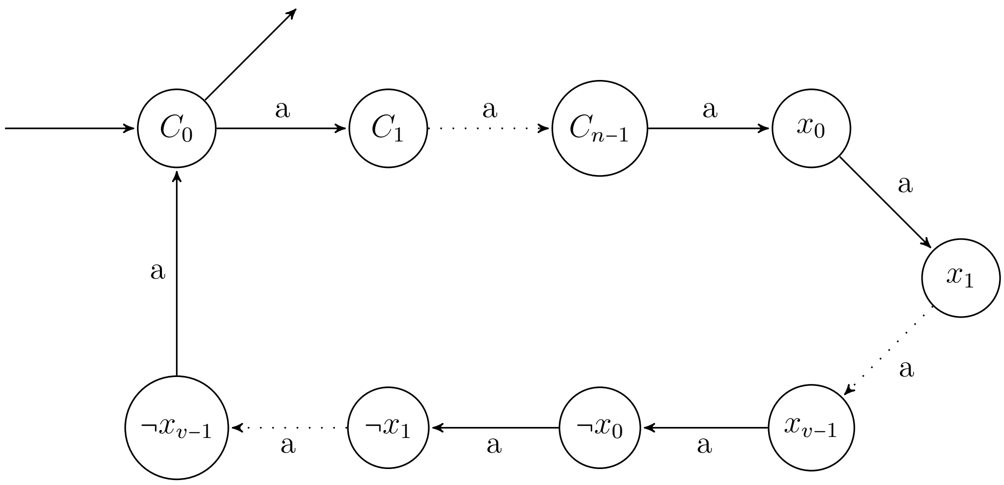
$$F = \bigwedge_{i=1}^{n-1} C_i, \quad \text{avec } x_0, \dots, x_{v-1} \text{ les variables de } F.$$

On désigne $x_0, \dots, x_{v-1}, \neg x_0, \dots, \neg x_{v-1}$ par $l_0, \dots, l_{2v-1}$. On va construire $(k(F), S(F), T(F))$ une instance de PSA tel que si $\exists \mathcal{A}$ à k états qui sépare S et T alors il existe une valuation δ de F tel que $\delta(F) = 1$.

Soit $k = n + 2v$, on va construire $S(F)$ et $T(F)$ sous forme d'union de sous-langages S_i et T_i.

L'objectif de toute la construction des S_i et T_i prennent en compte les propriétés que vérifient les différents objets du problème $SAT(F)$ et de donner l'automate les séparants lorsqu'il peut se construire. On construit alors $\mathcal{A}$ petit à petit.

1. Posons $S_1 = \{\epsilon, a^k\}$ et $T = \{a^i | 0 < i < k\}$. On construit l'automate $\mathcal{A}_1$ qui sépare S_1 et T_1 en nommant les sommets comme suit : $C_0, \dots, C_{n-1}, x_0, \dots, x_{v-1}, \neg x_0, \dots, \neg x_{v-1}$:



2. On veut maintenant que dans $\mathcal{A}_2$, on ait $C_j \xrightarrow{b} l_i$ lorsque l_i est une variable de C_j. C'est à dire que $\mathcal{A}_2$ est tel que $S_1 \subset \mathcal{A}_2$ et $(T_1 \cup T_2) \cap \mathcal{A}_2 = \varnothing$ avec $T_2 = \{a^i b a^j | 0 \geq i < net 0 \geq j < k\} \setminus \{a^i b a^{2v-j} | l_j \text{ terme de } C_i\}$.
3. On veut ensuite que $\forall i \in [|0, 2v-1|], l_i \xrightarrow{b} C_0$ ou $l_i \xrightarrow{b} \neg x_0$. C_0 et x_0 sont désormais vu comme VRAI et FAUX respectivement. C'est à dire que $\mathcal{A}_3$ sépare S_1 et $(T_1 \cup T_2 \cup T_3)$ où $T_3 = \{a^{n+j} b a^h | j \in [|0, 2v-1|], 0 < h < k \text{ et } h \neq v\}$.
4. On veut que les variables x_i et $\neg x_i$ ne soient jamais vraies en même temps, c'est à dire qu'il n'y a pas de succession de transitions dans $\mathcal{A}_4$ de la forme :

$$C_0 \xrightarrow{a^{n+j}} x_j \xrightarrow{b} C_0 \xrightarrow{a^{n+v+j}} \neg x_j \xrightarrow{b} C_0, \quad \forall j \in [|0, v-1|]$$

   C'est à dire que $\mathcal{A}_4$ sépare S_1 et $(T_1 \cup T_2 \cup T_3 \cup T_4)$ où $T_4 = \{a^{n+j} b a^{n+v+j} b | \forall j \in [|0, v-1|]\}$
5. Enfin, on veut que si $C_i \xrightarrow{b} l_j$ alors $l_j \xrightarrow{b} VRAI/(= C_0)$. C'est à dire que $\mathcal{A}_5$ sépare $(S_1 \cup S_2)$ et $(T_1 \cup \dots \cup T_4)$ où $S_2 = \{a^i bb | \forall i \in [|0, n-1|]\}$.

On a ainsi une application $F \mapsto (k(F), S(F), T(F))$ calculable en temps polynomial car S et T sont en $O(n+v)$. Il reste a bien vérifier que c'est bien une réduction de SAT à PSA.

Si l'automate $\mathcal{A}$ à k états peut bel et bien être construit comme plus haut, il sépare S et T et alors on pose la valuation $\delta : \forall i \in [|1, v-1|], \delta(x_i) = 1$ si $x_i \xrightarrow{b} C_0$, 0 sinon. Alors $\forall j$, la clause C_j est reliée à un littéral l_i tel que $l_i \xrightarrow{b} C_0$. Si l_i est une variable x, $\delta(x) = 1$ implique que C_j est satisfaite, si l_i est une variable $\neg x$, alors $\delta(\neg x) = 1 - \delta(x) = 1$ et implique encore que C_j est satisfaite. Donc $\delta(F) = 1$.

Remarque : Réciproquement, si $F \in SAT$, soit ϕ valuation telle que $\phi(F) = 1$. On note $t(C_i)$ le premier littéral de la classe C_i tel que $\phi(t(C_1)) = 1$. On construit alors l'automate déterministe $\mathcal{A} = (Q, \Sigma, \delta, C_0, C_0)$ avec $Q = \{C_0, \ldots, C_{n-1}, x_0, \ldots, x_{v-1}, \neg x_0, \ldots, \neg x_{v-1}\}$ et δ définie de la manière suivante :

$$\begin{array}{rcc}
\delta(q_i, a) & = & q_{i+1 \bmod k} \quad \forall 0 \geq i < k \\
\delta(C_i, b) & = & t(C_i) \quad \forall i \in [|0, n-1|] \\
\delta(l_i, b) & = & \delta(l_i) = C_0 \text{ si } \delta(l_i) = 1 \text{ pour } 0 \geq 1 < 2v, \quad \neg x_0 \text{ sinon}
\end{array}$$

On a alors $S_F \subset \mathcal{L}(\mathcal{A})$ et $T_F \cap \mathcal{L}(\mathcal{A}) = \varnothing$.