# Problème de séparation par automate

Référence : Le langage des machines.
Robert Floyd et Richard Biegel

2011-2012

Prérequis : *SAT est NP-complet.*

Alice et Bob jouent à un jeu :
- Alice pense à un langage $L$, et doit le faire deviner à Bob.
- Alice donne à Bob un ensemble de mots de $L$, $S$, et un ensemble de mot qui ne sont pas dans $L$, $T$.
- Bob doit trouver, à partir des exemples de $S$ et des contre-exemples de $T$ le langage $L$.

Le problème de séparation par automate consiste à chercher l'existence, pour des langages $S$ et $T$ finis et un entier $k$ donnés, un automate fini déterministe à $k$ états qui accepte tous les mots de $S$, et refuse tous les mots de $T$.

**Théorème 1**
*Le problème de séparation par automate (PSA) est NP-complet.*

*Démonstration.* Tout d'abord, PSA est de classe NP. En effet, on peut en temps polynomial construire de façon non déterministe un automate, et le tester sur les mots de $S$ et $T$.

Montrons donc que PSA est NP-dur. Nous allons pour cela réduire SAT à PSA : on suppose pouvoir résoudre PSA en temps polynomial, et on en déduit un algorothme de résolution de SAT en temps polynomial.

Soit donc $F$ une formule du calcul propositionel sous forme normale conjonctive. On note $C_0, \ldots, C_{m-1}$ les clauses de $F$, et $x_0, \ldots, x_{v-1}$ les variables de $F$.

Dans la suite on notera $\ell_0, \ldots, \ell_{2v-1}$ les littéraux $x_0, \ldots, x_{v-1}, \overline{x_0}, \ldots, \overline{x_{v-1}}$.

On pose aussi $k = 2v + m$.

On va construire deux ensembles $S$ et $T$ tels que si $P$ est un automate fini déterministe qui sépare $S$ et $T$, alors on peut déduire polynomialement de $P$ une valuation qui valide $F$.

Notre automate va en fait choisir, pour chaque clause de $F$, un littéral qui doit être mis à vrai, tout en s'assurant qu'un littéral et sa négation ne soit pas mis à vrai en même temps.

(1) On pose
  - $S_1 = \{\varepsilon, a^k\}$
  - $T_1 = \{a_i \mid 0 < i < k\}$.

  Comme $P$ a $k$ états, on les appelera dans l'ordre $C_0, \ldots, C_{m-1}, \ell_0, \ldots, \ell_{2v-1}$. $C_0$ est l'unique état initial, et l'unique état final.

(2) On veut que les arêtes issues d'un état $C_i$ et étiquetée par $b$ aillent vers un littéral de $C_i$. On pose alors :

$$T_2 = \{a^i b a^j \mid 0 \leqslant i < m,\ 0 \leqslant j < k\} \backslash \{a^i b a^{2v-j} \mid \ell_j \text{ est un littéral de } C_i\}.$$

**Lemme 2**
*Si $P$ rejette aussi les mots de $T_2$, alors les arêtes issues d'un état $C_i$ étiquetées par $b$ vont vers un littéral de $C_i$.*

*Démonstration.* Supposons qu'une arête étiquetée par $b$ aille de $C_i$ à $C_j$.
Soit alors $u = a^i ba^{k-j}$. Alors $u$ est accepté par $P$ et appartient à $T_2$, ce qui est impossible.
Donc les arêtes partant de $C_i$ étiquetée par $b$ vont nécessairement vers un $\ell_j$.
La lecture de $a^i b$ va de $C_0$ à $\ell_j$, et la lecture ensuite de $a^{2v-j}$ revient en $C_0$, et donc $a^i ba^{2v-j}$ est accepté par $P$, et donc $\ell_j$ est un littéral de $C_i$. $\Diamond$

(3) On veut que toutes les arêtes issues de $\ell_j$ étiquetées par $b$ aillent soit vers $C_0$ (on dit alors que $\ell_j$ est véridique), soit vers $\overline{x_0}$.
On pose alors

$$T_3 = \{a^{m+j} ba^h \mid 0 \leqslant j < 2v,\ 0 < h < k,\ h \neq v\}.$$

**Lemme 3**
*Un automate qui rejette aussi $T_3$ respecte la condition ci-dessus.*

*Démonstration.* Soit $P$ qui rejette aussi $T_3$. Supposons qu'il existe une arête étiquetée par $b$ qui va de $\ell_j$ à un état $q$ autre que $C_0$ ou $\overline{x_0}$.
Si $q = C_w,\ w \neq 0$, alors le mot $u = a^{m+j} b$ arrive en $C_w$, et donc le mot $a^{m+j} ba^{m-w+2v}$ est accepté $\rightarrow$ NON.
Si $q = \ell_w,\ w \neq v$, alors le mot $u = a^{m+j} b$ arrive en $\ell_w$, et donc $a^{m+j} ba^{2v-w}$ est accepté $\rightarrow$ NON. $\Diamond$

(4) On veut interdire que les littéraux $x_j$ et $\overline{x_j}$ soit véridiques en même temps. On pose donc

$$T_4 = \{a^{m+j} ba^{m+v+j} b \mid 0 \leqslant j < v\}.$$

**Lemme 4**
*Si $P$ refuse aussi $T_4$, alors $P$ satisfait l'hypothèse ci-dessus.*

*Démonstration.* Soit $P$ qui refuse aussi $T_4$. Supposons que $x_j \xrightarrow{b} C_0$ et $\overline{x_j} \xrightarrow{b} C_0$.
Alors $a^{m+j} b$ arrive en $C_0$, et $a^{m+v+j} b$ aussi, et donc la concaténation des deux est acceptée $\rightarrow$ NON. $\Diamond$

(5) Enfin, on veut que l'arête étiquetée par $b$ issue de $C_i$ (voir point (2)) aille vers un littéral véridique. On pose :

$$S_5 = \{a^i bb \mid 0 \leqslant i < m\}.$$

On prend alors $S = S_1 \cup S_5$ et $T = T_1 \cup T_2 \cup T_3 \cup T_4$.

Si $P$ à $k$ états accepte $S$ et rejette $T$, alors on construit la valuation :

$$\rho(x_i) = \begin{cases} \text{vrai} & \text{si } (x_j \xrightarrow{b} C_0) \text{ arête de } P \\ \text{faux} & \text{sinon} \end{cases}$$

Alors la valuation $\rho$ satisfait $F$.

$\square$