Huffman

Quentin Januel, Loïc Mohin et Anthony Villeneuve

2 février 2019

Table des matières

1 Définitions

1.1 Alphabet

On appelle Σ un alphabet dont les éléments sont appelés des lettres.
Un mot sur Σ est un n-uplet de lettres : $m = (a_1,\ a_2,\ ...,\ a_n)$.
L'ensemble des mots sur Σ est noté $\Sigma^* := \{m \in \Sigma^n,\ \forall n \in \mathbb{N}\}$.
Soit $m = (a_1,\ a_2,\ ...,\ a_n)$ un mot, on appelle longueur du mot m notée $|m|$ l'entier n.
Enfin, on note ε le mot vide (unique mot de longueur 0).

On peut munir Σ^* d'une loi de composition interne, la concaténation $+$:
$(a_1,\ ...,\ a_n) + (b_1,\ ...,\ b_n) = (a_1,\ ...,\ a_n,\ b_1,\ ...,\ b_n)$.
On observe alors que $(\Sigma^*,\ +)$ est un monoïde.

1.2 Arbre binaire

Soit E un ensemble, on dit que $A := (Q,\ T)$ est un arbre binaire sur E avec $Q \subset E$ et $T \subset E \times \mathbb{F}_2 \times E$ s'il respecte les 3 propriétés suivantes :

1. $\exists\ !\ r \in Q,\ \forall (x,\ b) \in Q \times \mathbb{F}_2,\ (x, b, r) \notin T$ (r est appelée racine de A notée $r(A)$),
2. $\forall x_2 \in Q \backslash \{r\},\ \exists\ !\ (x_1,\ b) \in Q \times \mathbb{F}_2,\ (x_1,\ b,\ x_2) \in T$,
3. $\forall (x_1,\ b) \in Q \times \mathbb{F}_2,\ \mathrm{card}(\{x_2 \in Q,\ (x_1,\ b,\ x_2) \in T\}) \leq 1$.

Les éléments de Q (notés $q(A)$) sont appelés les états et les éléments de T (notés $t(A)$) transitions.
Enfin, l'ensemble des arbres binaires sur E est noté $\mathcal{A}_E$.

1.3 Fonctions gauche et droite d'un arbre binaire

Soit E un ensemble, pour tout $b \in \mathbb{F}_2$, on pose

$$S_b := \{A \in \mathcal{A}_E,\ \exists\ x \in q(A),\ (r(A), b, x) \in t(A)\}$$

Cela correspond à l'ensemble des arbres binaires qui ont un sous arbre ou gauche ou droit selon la valeur de b.
Ensuite, $\forall A \in S_b$, soit $s_b(A)$ l'unique $x \in q(A)$ tel que $(r(A),\ b,\ x) \in t(A)$.
Finalement, on peut poser

$$\begin{array}{rccc} \theta_b : & S_b & \to & \mathcal{A}_E \\ & A & \mapsto & A_b \end{array}$$

avec A_b l'unique arbre binaire dont l'ensemble des transitions est aussi grand que possible tel que

$$r(A_b) = s_b(A) \tag{1}$$
$$t(A_b) \subset t(A) \tag{2}$$

Utilisons cette fonction pour définir $g := \theta_0$ et $d := \theta_1$.
Pour synthétiser, g et d sont deux fonctions qui rendent le sous arbre binaire respectivement gauche ou droite, définies uniquement si ce sous arbre existe.

1.4 Arbre de Huffman

On appelle arbre de Huffman un arbre binaire A sur $\mathbb{N}$ qui respecte les propriétés suivantes :
- $n(A) = n(g(A)) + n(d(A))$,
- $g(A)$ et $d(A)$ sont tous deux des arbres de Huffman,
- ... (il manque un truc ici, à faire !).

On note $\mathcal{H}$ l'ensemble des arbres de Huffman et on a $\mathcal{H} \subset \mathcal{A}_{\mathbb{N}}$.

2 Classes d'équivalence d'arbres de Huffman

Intuitivement, on aimerait définir des classes d'équivalences des arbres de Huffman selon la longueur de la chaine de bits qu'ils encodent. Tâchons donc d'abord de pouvoir évaluer cette valeur à l'aide d'une fonction :

$$\begin{array}{rccc} \omega : & \mathcal{H} & \to & \mathbb{N} \\ & A & \mapsto & \begin{cases} 0 \text{ si } A = A_0 \\ n(A) + \omega(g(A)) + \omega(d(A)) \text{ sinon} \end{cases} \end{array}$$

Nous pouvons à présent définir $\mathcal{R}$ notre relation d'équivalence. Soient $A, B \in \mathcal{H}$, on a alors

$$A \mathcal{R} B \iff \omega(A) = \omega(B)$$

On dénote également $\overline{\mathcal{H}} := \mathcal{H}/\mathcal{R}$ l'ensemble quotient de $\mathcal{H}$ par $\mathcal{R}$.