Optimalitatea Algoritmului Huffman

Pentru un arbore binar strict cu ponderi, numim in continuare Cost lungimea externa ponderata

$$Cost(T) = \Sigma_{e \in \mathcal{E}} w(e) h_T(e),$$

unde $h_T(e)$ este adancimea frunzei e in T si este egala cu lungimea in arce a drumului de la radacina la frunza e.

Un arbore de codificare T este **optim** daca pt. orice alt arbore de codificare T' peste acelasi alfabet avem relatia

$$Cost(T) = \Sigma_{e \in \mathcal{E}} w(e) h_T(e) \leq \Sigma_{e \in \mathcal{E}} w(e) h_{T'}(e) = Cost(T').$$

Lemma 0.1 *Fie T un arbore de codificare optim peste $\mathcal{E}$. Daca $w(e) < w(e')$ atunci $h_T(e) \geq h_T(e')$.*

Demonstratie: Sa presupunem prin absurd ca avem o pereche de frunze care indeplinesc simultan conditiile $w(e) < w(e')$ si $h_T(e) < h_T(e')$. Schimbam intre ele frunzele e si e' obtinand un arbore binar de codificare T' de cost

$$Cost(T') = Cost(T) - w(e)h_T(e) - w(e')h_T(e') + w(e)h_T(e') + w(e')h_T(e) =$$
$$= Cost(T) - (w(e) - w(e'))(h_T(e) - h_T(e')) < Cost(T),$$

ceea ce contrazice optimalitatea lui T.

Lemma 0.2 *Fie w_1 si w_2 cele mai mici ponderi, corespunzatoare frunzelor e_1 si e_2. Atunci, exista un arbore optim in care e_1 si e_2 sunt frati.*

Demonstratie: Exercitiu!

Theorem 0.1 *Algoritmul lui Huffman construieste un arbore binar de codificare optim.*

Demonstratie: Prin inductie dupa $n = |\mathcal{E}|$. Pentru $n \leq 2$ avem un singur arbore posibil si teorema e evident adevarata. Sa presupunem ca $n \geq 3$. Notam cu T_H arborele construit cu algoritmul lui Huffman pentru ponderile $w_1 \leq w_2 \leq \cdots \leq w_n$. Algoritmul leaga intre ele frunzele e_1 si e_2 corespunzatoare ponderilor w_1 si w_2 si creeaza un arbore cu ponderea $w_1 + w_2$. Fie T'_H arborele construit de algoritmul lu Huffman din ponderile $w_1 + w_2, w_3, \cdots, w_n$. Avem relatia

$$Cost(T_H) = Cost(T'_H) + w_1 + w_2.$$

Conform ipotezei de inductie, T'_H este arbore optim pentru $n - 1$ frunze cu ponderile $w_1 + w_2, w_3, \cdots, w_n$.

Fie T_{opt} un arbore de codificare optim ce satisface Lema 2, adica frunzele e_1 si e_2 sunt frati in T_{opt}. Fie T' arborele obtinut din T_{opt} inlocuind frunzele e_1 si e_2 si pe tatal lor cu o singura frunza de pondere $w_1 + w_2$. Apicand ipoteze de inductie $Cost(T'_H) \leq Cost(T')$, avem

$$Cost(T_{opt}) = Cost(T') + w_1 + w_2 \geq Cost(T'_H) + w_1 + w_2 = Cost(T_H).$$

Rezulta deci

$$Cost(T_{opt}) = Cost(T_H).$$