Zadanie 4 z listy 2 - "Kompresja Danych"

Łukasz Klasiński

30 marca 2020

Zadanie 4

Pokaż, że kod Shannona jest kodem prefiksowym, ale nie jest optymalnym kodem prefiksowym.

Rozwiązanie

Przypomnijmy najpierw konstrukcję kodów Shannona:

- $p_1 \geq \cdots \geq p_n$ - posortowane prawdopodobieństwa symboli $s_1 \dots s_n$
- $F_i = p_1 + p_2 + \cdots + p_{i-1}$
- $l_i = \lceil log(1/p_i) \rceil$ - liczba bitów liczby F_i ('po przecinku'), które użyte zostaną do reprezentacji symbolu s_i

D-d tego, że kody Shannona są prefiksowe

Zauważmy najpierw, że z konstrukcji l_i dochodzimy do następującej nierówności:

$$2^{-l_i} \leq p_i < 2^{(-(l_i-1))}$$

Weźmy dowolne F_j oraz F_i takie, że $j > i$. Wtedy F_j różni się od F_i o co najmniej 2^{-l_i} ponieważ $F_i = F_j - (p_i + \dots)$, a $p_i \geq 2^{-l_i}$. Oznacza to, że F_j różni się od F_i w przynajmniej jednym bicie zawierającym się w pierwszych l_i bitach rozszerzenia binarnego F_i. Zatem kod dla F_j, który ma długość $l_j \geq l_i$, różni się od kodu F_i w co najmniej jednym miejscu (pierwsze l_i bitów), więc kod F_i nie może być jego prefiksem. Z dowolności wyboru wynika, że żadne słowo kodowe nie jest prefiksem innego.

$\square$

D-d, że kod Shannona nie jest optymalnym kodem prefiksowym.

Załóżmy, że kod shannona jest optymalny. Weźmy dane:

- symbole - $\{A, B\}$
- prawdopodobieństwa - $\{p_A = \frac{63}{64}; p_B = \frac{1}{64}\}$
- sort - $\{p_0 = \frac{63}{64}, p_1 = \frac{1}{64}\}$

Algorytm Shannona przydzieli następujące kody prefiksowe:

- A = 0, bo $\lceil log(\frac{1}{p_A}) \rceil = 1, F_0 = 0.0$
- B = 111111, bo $\lceil log(\frac{1}{p_B}) \rceil = 6$, a $F_1 = \frac{63}{64} = 0.111111$

Zatem kod prefiksowy Shannona nie jest optymalny, bo optymalne byłoby przypisanie A i B po jednym, różnym bicie - otrzymujemy sprzeczność.

$\square$