Matematyka dyskretna, WMat 2024

Lista 3: Kodowanie

1. Które z tych kodowań są jednoznacznie dekodowalne? Które są przedrostkowe?

(a) $C_a = [\texttt{1}, \texttt{101}]$,

(b) $C_b = [\texttt{0}, \texttt{101}]$,

(c) $C_c = [\texttt{111}, \texttt{101}, \texttt{110}, \texttt{011}, \texttt{01}]$,

(d) $C_d = [\texttt{0}, \texttt{10}, \texttt{110}, \texttt{111}]$,

(e) $C_e = [\texttt{0}, \texttt{01}, \texttt{011}, \texttt{111}]$,

(f) $C_f = [\texttt{00}, \texttt{22}, \texttt{100}, \texttt{201}, \texttt{212}, \texttt{0110}, \texttt{0112}]$.

Dla kodowań przedrostkowych narysuj ich drzewo decyzyjne.

2. Niech dany będzie kodowanie 4-symbolowe, które ma 3 wyrazy długości 1 oraz 6 wyrazów długości 2. Czy może ono być jednoznacznie dekodowalne?

3. Mając tabelę

Symbol	$p(x)$	$q(x)$	$c_1(x)$	$c_2(x)$
a	1/2	1/2	0	0
b	1/4	1/8	10	100
c	1/8	1/8	110	101
d	1/16	1/8	1110	110
e	1/16	1/8	1111	111

Oblicz średnią długość kodowania C_1 przy rozkładzie p oraz C_2 przy rozkładzie q. Sprawdź, że są optymalne. Następnie oblicz, ile tracimy używając C_2 przy rozkładzie p oraz C_1 przy rozkładzie q. Porównaj ten wynik z wartościami entropii relatywnych $D(p\|q)$ oraz $D(q\|p)$ oraz wyciągnij wnioski.

4. Oblicz średnią długość kodowania, w którym liczby od 1 do 100, z których każda ma równe prawdopodobieństwo, przedstawiono w systemie:

(a) jedynkowym,

(b) dwójkowym,

(c) trójkowym.

5. Znajdź optymalne jednoznacznie dekodowalne kodowanie binarne dla zmiennej o 7 symbolach mających równe prawdopodobieństwa.

6. Pokaż, że jeżeli zmienna ma rozkład o prawdopodobieństwach równych ujemnych potęgom liczby 2, to można znaleźć kodowanie o średniej długości równej entropii.

7. Które z tych kodowań nie mogą być kodowaniami Huffmana?

(a) $C_a = [\mathtt{0}, \mathtt{10}, \mathtt{11}]$,

(b) $C_b = [\mathtt{00}, \mathtt{01}, \mathtt{10}, \mathtt{11}]$,

(c) $C_c = [\mathtt{01}, \mathtt{10}]$.

8. Skonstruuj kodowanie Huffmana dla zmiennej X o alfabecie $\mathcal{A} = [\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}, \mathtt{e}, \mathtt{f}, \mathtt{g}]$ i rozkładzie $p = [0.01, 0.24, 0.05, 0.2, 0.47, 0.01, 0.02]$.

9. Skonstruuj kodownie Huffmana dla $XX, XXX, XXXX$, gdzie X ma 2 symbole o prawdopodobieństwach $p_X = [3/4, 1/4]$. Porównaj średnią długość wyrazów tego kodowania do entropii $H(XX), H(XXX), H(XXXX)$.

10. Podaj przykład, w którym różnica między entropią a średnią długością kodu Huffmana jest dowolnie bliska górnej granicy, tzn. kodowanie Huffmana jest tak nieefektywne, jak to tylko możliwe.

11. Załóżmy, że grasz w zabawę„20 pytań" z bogiem, który zawsze zadaje optymalne pytania, dzięki którym średnio potrzebuje ich jak najmniej tylko możliwe. Do zidentyfikowania obiektu bóg potrzebuje średnio 38.5 pytania. Oszacuj minimalną liczbę identyfikowalnych obiektów we Wszechświecie.