

## Spis treści

<b>1</b>	<b>Wstęp</b>	<b>1</b>
1.1	"Prawo" Kompresji bezstratnej . . . . .	1
1.2	Kodowanie . . . . .	1
1.2.1	Modelowanie danych . . . . .	1
1.2.2	Średnia długość kodu . . . . .	1
1.2.3	Jednoznaczna dekodowalność . . . . .	2
1.2.4	Nierówność Krafta . . . . .	2
1.2.5	Kod natychmiastowy . . . . .	2
1.2.6	Kod prefiksowy . . . . .	2
1.2.7	Kod Huffmana . . . . .	2
<b>2</b>	<b>Teoria informacji</b>	<b>2</b>
2.1	Entropia . . . . .	2

## 1 Wstęp

Wyróżniamy dwa rodzaje kompresji. W kompresji stratnej dopuszczalny jest pewien stopień straty informacji wejściowej. W kompresji bezstratnej nie jest to dopuszczalne.

### 1.1 "Prawo" Kompresji bezstratnej

Nie istnieje algorytm, który potrafi zmniejszyć rozmiar dowolnych danych

- Kompresja bezstratna musi być bijekcją
- Dowolne dane przyjmują postać ciągu bitów długości  $n$ . Jest  $2^n$  takich ciągów.
- Danych krótszych niż  $n$ , np.: o jeden jest  $2^{n-1}$
- Nie da się stworzyć bijekcji z zbioru o mocy  $2^n$  do zbioru o mocy  $2^{n-1}$

Wniosek jest taki, że koniecznym jest konstruowanie kompresji bezstratnej na podzbiorach danych, takich jak np.: obrazów, dźwięków, tekstów.

### 1.2 Kodowanie

Kodowanie to przyporządkowanie elementom jakiegoś alfabetu ciągu binarnych. Przykładami kodowania są: ASCII, UTF-8 oraz inne. Typowym jest konstruowanie kodowania pod konkretny zestaw danych, optymalizując je pod kątem częstości występowania poszczególnych elementów.

#### 1.2.1 Modelowanie danych

Rozważmy ciąg:  $a_n = 9, 11, 11, 11, 14, 13, 15, 17, 16, 17, 20, 21$ .  $\max(a_n) = 21$  stąd koniecznym jest 5 bitów na element. Ale jeśli wykorzystamy wzór  $e_n = a_n - n + 8$  do stworzenia nowego ciągu, to ten ciąg przyjmuje postać:  $0, 1, 0, -1, 1, -1, 0, 1, -1, -1, 1, 1$ . Teraz wystarczy tylko 2 bity na zakodowanie elementu.

#### 1.2.2 Średnia długość kodu

$$I = \sum_{i=1}^n p_i \cdot l_i$$

gdzie  $p_i$  to prawdopodobieństwo wystąpienia elementu  $i$ , a  $l_i$  to długość kodu dla elementu  $i$ .

### 1.2.3 Jednoznaczna dekodowalność

Jeśli dla dowolnego ciągu znaków istnieje tylko jedno jego rozkodowanie to kod jest jednoznacznie dekodowalny. Aby sprawdzić czy kod jest jednoznacznie dekodowalny, należy zastosować następujący algorytm.

1. Stwórz pustą listę
2. Dla każdej pary słów kodowych sprawdź czy jedno jest prefiksem drugiego. Jeśli tak, dodaj sufiks drugiego słowa do listy, jeśli już go tam nie ma.
3. Jeśli na liście jest słowo kodowe, to kod nie jest jednoznacznie dekodowalny.

### 1.2.4 Nierówność Krafta

Jeżeli  $\mathcal{C}$  jest kodem jednoznacznie dekodowalnym z  $n$  słowami to:

$$K(\mathcal{C}) = \sum_{i=1}^n 2^{-l_i} \leq 1$$

Jest to warunek konieczny bycia kodem jednoznacznie dekodowalnym.

### 1.2.5 Kod natychmiastowy

Jest kodem pozwalającym stwierdzić w którym miejscu zakończone jest słowo kodowe w momencie odczytania ostatniej litery.

### 1.2.6 Kod prefiksowy

Kod w którym żadne słowo kodowe nie jest prefiksem innego słowa kodowego. Wszystkie kody prefiksowe są jednoznacznie dekodowalne.

### 1.2.7 Kod Huffmana

Kod Huffmana to kod prefiksowy o minimalnej średniej długości kodu.

## 2 Teoria informacji

Miarą informacji, którą niesie ze sobą zdarzenie  $A$  jest:

$$I(A) = -\log_x P(A)$$

gdzie  $x$  to baza systemu liczbowego. Jeśli miarą informacji jest bit to  $x = 2$ . Jeśli zdarzenia  $A$  i  $B$  są niezależne to:

$$I(AB) = I(A) + I(B)$$

### 2.1 Entropia

Dla źródła informacji  $X$ , z zbiorem wiadomości (zdarzeń)  $A_1, \dots, A_n$ , gdzie  $P(A_i)$  to prawdopodobieństwo wystąpienia zdarzenia  $A_i$  to entropia źródła to:

$$H(X) = \sum_{i=1}^n P(A_i) I(A_i)$$

Entropia to miara średniej informacji przekazywanej przez źródło. Kody jednoznacznie dekodowalne w modelu z niezależnymi wystąpieniami symboli muszą mieć średnią długość co najmniej równą entropii.