Politechnika Wrocławska

Struktury danych i złożoność obliczeniowa

Prof. dr hab. inż. Jan Magott

Kodowanie danych wejściowych

Sposób kodowania danych wejściowych ma wpływ na rozmiar problemu (długość danych wejściowych)

Czy każdy z następujących sposobów kodowania liczb:

- Dziesiętny,
- Binarny,
- Jedynkowy

jest „właściwy” ?

Alfabet jest skończonym zbiorem symboli.

Alfabet oznaczamy symbolem Σ .

Przykłady

Alfabet dziesiętny $\Sigma = \{0,1,2,\ldots,8,9,\sqcup\}$, gdzie $\sqcup$ jest separatorem.

Alfabet dwójkowy $\Sigma = \{0,1,\sqcup\}$.

Alfabet jedynkowy $\Sigma = \{1,\sqcup\}$.

Słowem alfabetu Σ nazywamy skończony ciąg symboli alfabetu $\Sigma\backslash\{\sqcup\}$.

Szczególnym przypadkiem jest słowo puste czyli nie zawierające żadnego symbolu

Przykłady

Słowa dziesiętne: 536, 1429, 374502

Słowa dwójkowe: 10, 0, 1001011

Słowa jedynkowe: 111, 11111, 1, 11111111

Słownikiem alfabetu $\sum$ jest zbiór wszystkich słów alfabetu $\sum\backslash\{\sqcup\}$.

Przykłady

Słownik dziesiętny:

{0,1,2,...,9,10,11,...,99,100, 101, ...}

Słownik dwójkowy:

{0,1,10,11,100,101, ...}

Słownik jedynkowy:

{1,11,111,1111,...}

Językiem alfabetu $\sum$ jest skończony ciąg słów tego alfabetu oddzielonych separatorami ␣ .

Przykłady

Język dziesiętny: 402␣31␣58

Język dwójkowy: 1011␣101111␣11␣1000

Język jedynkowy: 11␣1111␣1␣111␣111

Reguła kodowania

1. Kolejność danych musi być określona
2. Każda dana jest reprezentowana jednym słowem alfabetu $\sum$
3. Musi istnieć metoda wyodrębniania poszczególnych danych
4. Dane o równych wartościach kodowane są tym samym słowem

Instancja I problemu podziału zbioru

Dane:

- $X = \{x_1, \dots, x_i, \dots, x_k\} = \{3,11,7,23,18\}$ - zbiór pięciu elementów,
- $B = 32, \quad \sum_{i=1}^{5} x_i = 2 \cdot B.$

Pytanie:

Czy istnieje podzbiór $X_1 \subset X$ taki, że $\sum_{x_i \in X_1} x_i = B$?

Zakodujmy dane instancji I w następujący sposób:

$$x_1 \sqcup x_2 \sqcup x_3 \sqcup x_4 \sqcup x_5 \sqcup B \sqcup \ \sqcup$$

Kodowanie w alfabecie dziesiętnym

$$3 \sqcup 11 \sqcup 7 \sqcup 23 \sqcup 18 \sqcup 32 \sqcup \ \sqcup$$

Długość łańcucha kodującego instancję I wynosi

$$N_{10}(I) = 17$$

Zakodujmy dane instancji I w następujący sposób:

$$x_1 \sqcup x_2 \sqcup x_3 \sqcup x_4 \sqcup x_5 \sqcup B \sqcup \ \sqcup$$

Kodowanie w alfabecie dwójkowym

$$11 \sqcup 1011 \sqcup 111 \sqcup 10111 \sqcup 10010 \sqcup 100000 \sqcup$$

Długość łańcucha kodującego instancję I wynosi

$$N_2(I) = 31$$

Zakodujmy dane instancji I w następujący sposób:

$$x_1 \sqcup x_2 \sqcup x_3 \sqcup x_4 \sqcup x_5 \sqcup B \ \sqcup \ \sqcup$$

Kodowanie w alfabecie jedynkowym

$111\sqcup 11111111111\sqcup 1111111\sqcup \ldots .111111111\sqcup \ \sqcup$

Długość łańcucha kodującego instancję I wynosi

$$N_1(I) = 100$$

Problem podziału zbioru

Dane:

- $X = \{x_1, \dots, x_i, \dots, x_k\}$ - zbiór k elementów $x_i \in N_+$, gdzie $N_+ = \{1,2, \dots\}$,
- $B \in N_+$,
- $\sum_{i=1}^{k} x_i = 2B$.

Pytanie:

Czy istnieje podzbiór $X_1 \subset X$ taki, że

$$\sum_{x_i \in X_1} x_i = B \text{ ?}$$

Zakodujmy dane Problemu podziału w następujący sposób:

$$x_1 \sqcup x_2 \sqcup \dots x_n \sqcup B \sqcup \ \sqcup$$

Długość łańcucha kodującego w alfabecie dziesiętnym

$$N_{10}(I) = \lceil log x_1 \rceil + 1 + \lceil log x_2 \rceil + 1 + \dots + \lceil log x_n \rceil + 1 + \lceil log B \rceil + 2 = \sum_{i=1}^{n} \lceil log x_i \rceil + n + \lceil log B \rceil + 2$$

$$\leq (n+1) \cdot max_{j \in \{\overline{1,n}\}} \{\lceil log x_i \rceil + 1, \lceil log B \rceil\} + 2$$

$$\approx O(n \cdot max_{j \in \{\overline{1,n}\}} \{\lceil log x_i \rceil + 1, \lceil log B \rceil\})$$

$\lceil x \rceil$ - najmniejsza liczba całkowita nie mniejsza niż x

Zakodujmy dane Problemu podziału w następujący sposób:

$$x_1 \sqcup x_2 \sqcup \ldots x_n \sqcup B \sqcup \;\; \sqcup$$

Długość łańcucha kodującego w alfabecie dwójkowym

$$N_2(I) = \lceil log_2 x_1 \rceil + 1 + \lceil log_2 x_2 \rceil + 1 + \ldots + \lceil log_2 x_n \rceil + 1 + \lceil log_2 B \rceil + 2 = \sum_{i=1}^{n} \lceil log_2 x_i \rceil + n + \lceil log_2 B \rceil + 2 \leq$$

$$\leq (n+1) \cdot max_{j \in \{\overline{1,n}\}} \{\lceil log_2 x_i \rceil + 1, \lceil log_2 B \rceil\} + 2$$

$$\approx O(n \cdot max_{j \in \{\overline{1,n}\}} \{\lceil log_2 x_i \rceil + 1, \lceil log_2 B \rceil\}$$

$\lceil x \rceil$ - najmniejsza liczba całkowita nie mniejsza niż x

Zakodujmy dane Problemu podziału w następujący sposób:

$$x_1 \sqcup x_2 \sqcup \ldots x_n \sqcup B \ \sqcup \ \sqcup$$

Długość łańcucha kodującego w alfabecie jedynkowym

$$N_1 = x_1 + 1 + x_2 + 1 + \cdots + x_n + 1 + B + 2 = \sum_{i=1}^{n} x_i + n + B + 2$$

$$x_i \approx 10^{\ log_{10} x_i}$$

Powyższej funkcji nie można ograniczyć wielomianem.

„Rozsądna i zwarta" reguła kodowania to reguła, która nie powoduje wykładniczego wzrostu rozmiaru danych wejściowych kodowanej instancji w stosunku do innych reguł kodowania.

„Rozsądne i zwarte" reguły kodowania instancji $\boldsymbol{I}$ dające łańcuchy o długościach $N_n(I)$, $N_m(I)$ spełniają relacje:

$N_n(I) \leq p(N_m(I))$ i $N_m(I) \leq p'(N_n(I))$ dla wielomianów p , p'.

Jeśli ten wymóg nie jest spełniony i np. $N_n(I) \leq k^{N_m(I)}$ dla pewnej stałej $k > 1$, to pierwsza jest „nierozsądna" (powoduje wykładniczy wzrost rozmiaru).

Zatem odrzucamy kodowanie danych wejściowych w alfabecie jedynkowym.