

Złożoność informacyjna Kołmogorowa

przygotował: Janusz Bednarski
12-05-2004

Problem 1: Najmniejsza liczba naturalna, której nie można zdefiniować w języku polskim przy pomocy mniej niż 1000 symboli.

Stawiając ten problem zdefiniowaliśmy już tę liczbę.

Problem 2: Co to znaczy, że ciąg jest „losowy”?

przykład: Rzucamy 100 razy monetą, wynik zapisujemy w postaci $(0,1)$.

Wynikiem jest:

0000.....0 – nieprawdopodobny wynik

0101.....01 – bardziej prawdopodobny, ale jakiś dziwny

0110..... – wygląda na losowy, ale ...

Kryterium, które przybliża nas do losowości to iż każde słowo $w \in \{0, 1\}^n$ wystąpi z tą samą gęstością.

Ciąg Champernowna jest to ciąg, który powstaje poprzez zapisanie kolejnych liczb naturalnych w sposób taki, iż wypisywanych liczb nie oddzielamy przecinkami.

przykład: 1234567891011121314151617...

Ciąg Champernowna spełnia powyższe kryterium, a pomimo to trudno nam go uznać za ciąg losowy.

Nieformalnie, ciąg X jest losowy (random) w sensie Kołmogorowa jeśli najkrótszy program, który go generuje ma długość $\geq |X|$ (np `write(' X ')`).

Problem 3: W jakim sensie słowo 011010110111001 jest bardziej skomplikowane niż słowo 01010101010101? Ponieważ złożoność obliczeniowa zajmuje się nieskończonymi językami, a nie skończonymi słowami, więc na powyższe pytanie nie da odpowiedzi. Istnieje ważna teoria, złożoność opisowa, której podstawy stworzył A.N. Kołmogorow, a która zajmuje się takimi pytaniami.

Maszyna Turinga (przypomnienie):

Przypomnijmy podstawowe pojęcia maszyny Turinga oraz określmy jej podstawowe moduły:

- a) urządzenie operacyjne może przyjmować jeden stan spośród skończonej ich liczby $Q = \{q_0, q_1, \dots, q_n\}$, w zbiorze Q zaś są wydzielone początkowe stany q_1 oraz końcowe q_0 ,*
- b) nieskończona taśma jest podzielona na komórki, w każdej z których może być nie więcej niż jeden symbol skończonego alfabetu,*
- c) głowica może dokonywać odczytu oraz zapisu symboli alfabetu oraz przesuwając taśmę na prawo lub na lewo o jedną komórkę,*

W procesie sekwencyjnego wykonywania ustalonych zasad maszyna Turinga może zmieniać zawartości komórek na taśmie przesuwając ją lub przechodzić do innego stanu. Podczas przejścia do stanu końcowego q_0 , praca maszyny kończy się. Oczywiście jest, że podczas pracy maszyny Turinga odbywa się przekształcanie danych początkowych zapisanych na taśmie, przy czym nie musi nastąpić zatrzymanie maszyny – dopuszczalne jest jej „zapętlenie”.

Uniwersalna maszyna Turinga:

Uniwersalna maszyna Turinga $M^{(u)}$ jest to maszyna, która dostarczone jej dane interpretuje jako opis innej maszyny Turinga, złączony z opisem słowa wejściowego dla tej maszyny.

Przyjmijmy dla każdej maszyny alfabet 0, 1, B.

Kodowanie: 0, 1, B ; L, R, Z

Odpowiednie kody będą miały swoje odpowiedniki:

0 00 000 ; 0 00 000

Kod instrukcji:

$q_2, 1 \rightarrow q_5, 0, L$

00100100000101011 – przykładowy kod

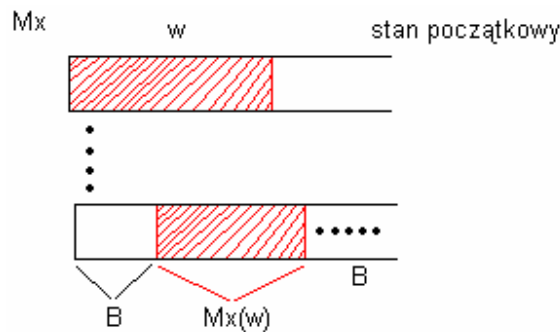
Kod maszyny:

111...11...11...111

M_x - maszyny o kodzie x

$M_x(w)$, gdzie w jest to słowo zerojedynkowe

$$M_x = \left\{ \begin{array}{l} \text{wynik obliczenia} \\ \uparrow \\ \text{–maszyna nie zakończyła działania} \end{array} \right\}$$



Maszyna uniwersalna $M^{(u)}(xw) = M_x(w)$ - jeśli x jest poprawnym kodem.

Istnieją dwie definicje. **Definicja I (prostsza):**

Określimy $K(w) =_{\text{def}} \min \{ |x| : M^{(u)}(x) = w \}$ - to nie jest zbiór pusty.

$$K(w|u) =_{\text{def}} \min \{ |x| : M^{(u)}(xu) = w \}$$

$$K(w) = K(w|\varepsilon)$$

Ogólnie dla dowolnej maszyny M , określimy:

$$K_M(w|u) = \min \{ |x| : M(xu) = w \}$$

FAKT:

Istnieje taka stała C_M , że $K(w|u) \leq K_M(w|u) + C_M$

Dowód:

$\langle M \rangle$ - kod maszyny M .

$$\text{Jeśli } M(xu) = w \Rightarrow M^{(u)}(\underbrace{\langle M \rangle xu}_{x'}) = w$$

Jeśli dobierzemy x tak, że $|x| = K_M(w|u)$, to wtedy złożoność

$$K(w|u) \leq |x'| = \underbrace{|x|}_{K_M(w|u)} + \underbrace{|\langle M \rangle|}_C$$

Uwaga: (dla złożoności warunkowej)

Liczba ciągów x , że $|x| < |w|$ jest $1 + 2 + 2 + \dots + 2^{|w|-1} = 2^{|w|} - 1$, więc istnieje ciąg w , że $K(w) \geq w$. Takie w nazywamy losowym.

Liczba x , że $|x| < |w| - 1$ jest $2^{|w|} - 1$ więc $2^{|w|} - 2^{|w|-1} + 1 > \frac{1}{2} 2^{|w|}$; ponad połowa słów w o danej długości spełnia $K(w) \geq |w| - 1$

TWIERDZENIE:

Istnieje taka stała d , że $K(w) \leq |w| + d$

Dowód:

Niech $M^{(C)}(z) = z$ - taka maszyna kopiująca.

Wtedy złożoność przy tej maszynie $K_{M^{(C)}} \leq |w|$

Z poprzedniego faktu:

$$K(w) \leq \underbrace{K_{M^{(C)}}(w)}_{\leq |w|} + C$$

TWIERDZENIE:

Funkcja $K: \{0,1\}^* \rightarrow \mathbb{N}$ nie jest obliczalna. Oznacza to iż nie istnieje maszyna Turinga, która ją oblicza.

Dowód (nie wprost):

Rozważmy funkcję $n \mapsto$ pierwszy (w porządku leksykalnym) ciąg w , że $K(w) \geq n$.

Z przypuszczenia ta funkcja jest obliczalna przez maszynę $\tilde{M}$ niech $w_n = \tilde{M}(n)$.

$$K_{\tilde{M}}(w_n) = (\log_2 n) + 1$$

$$n \leq K(w_n) \leq K_{\tilde{M}}(w_n) + C_{\tilde{M}} = (\log_2 n) + 1 + C_{\tilde{M}} - \text{a to jest sprzeczność!}$$