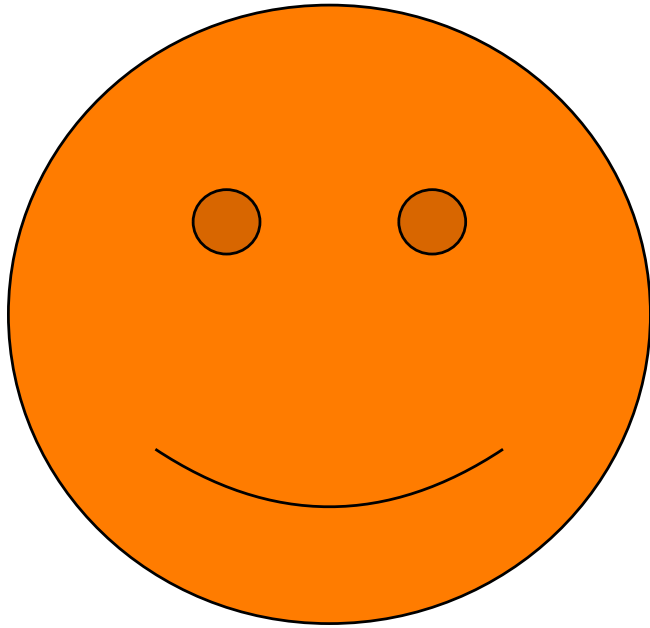


A digitális számítás elmélete

11. előadás május 3.

- Információtartalom algoritmikus alapokon
- Kolmogorov bonyolultság
- Kolmogorov véletlen sorozatok

Információtartalom – intuitív példák



Információtartalom – intuitív példák

Melyik sorozatnak a legtöbb ill. legkevesebb az információtartalma?

- 1111111111111111
- 1001011110110101
- 1010101010101010

Információtartalom – intuitív példák

Adjuk meg röviden az alábbi számokat:

- 8192 2^{13} 13=1101₂
+..

- 832040 F(30) F(n) = F(n-1) + F(n-2)

$$\frac{\Pi}{2} = \frac{2}{1} \cdot \frac{2}{3} \cdot \frac{4}{3} \cdot \frac{4}{5} \cdot \frac{6}{5} \cdot \frac{6}{7} \cdot \dots \cdot \frac{2n}{2n-1} \cdot \frac{2n}{2n+1} \dots$$

- 1415926535 8979323846

$$\frac{\Pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \dots + \frac{(-1)^n}{2n+1} + \dots$$

Információtartalom – intuitív példák

Rövid megadás

- természetes nyelven: értelmezési probléma, pl:
 - max. 100 írógép karakterrel megadható számok
 - ezen legnagyobbikának rákövetkezője?
- binárisan: jó, hossz azonos a bin. szám hosszával
- **Turing gépet használva**

Összenyomhatóság

Def: $I = \{0, 1\}$, M egy olyan TG, melynek bemeneti és szalag abc-je I .

M az $f_M: I^* \rightarrow I^*$ parciális függvényt számolja ki.

Ekkor egy I^* -beli x szó esetén

$$C_M(x) = \begin{cases} \min \{ |y| : y \in I^*, f_M(y) = x \} & \text{ha van ilyen } y \\ \infty & \text{különben} \end{cases}$$

$C_M(x)$ azt adja meg, hogy M milyen mértékben nyomja össze x -et.

Megj: $C_M(x)$ függ M -től! $C_{M_1}(x) = \infty$ $C_{M_2}(x) = 0$ alkalmas M_i -re.

Összenyomhatóság

Tétel: (Invariancia tétel)

Legyen U egy univerzális TG. Ekkor tetszőleges M TG-hez létezik $c_M \in \mathbb{Z}^+$ szám, hogy minden I^* -beli x szó esetén

$$C_U(x) \leq C_M(x) + c_M.$$

Biz: Legyen y egy legrövidebb szó, melyre $f_M(y) = x$.

Legyen M kódja $w \in I^*$.

Ekkor U a $w\#y$ bemenettel x -et adja meg, azaz $f_U(w\#y) = x$.

$$\text{Tehát } C_U(x) \leq |w\#y| = |w\#| + |y| = |w\#| + C_M(x) = \underbrace{c_M}_{\text{ }} + C_M(x)$$

Összenyomhatóság

Megj: (Invariancia tétel következménye)

Legyen U_1 és U_2 univerzális TG-k. Ekkor létezik olyan

$c = c_{U_1, U_2} \in \mathbb{Z}^+$ szám, hogy minden I^* -beli x szó esetén

$$| C_{U_1}(x) - C_{U_2}(x) | \leq c.$$

Biz: Az invariancia tételt $U = U_1$, $M = U_2$ és $U = U_2$, $M = U_1$

szereposztással felírva

$$C_{U_1}(x) - C_{U_2}(x) \leq c_{U_2} \text{ és } C_{U_2}(x) - C_{U_1}(x) \leq c_{U_1}$$

$$\text{Tehát } | C_{U_1}(x) - C_{U_2}(x) | \leq c = \max \{ c_{U_1}, c_{U_2} \}.$$

Kolmogorov bonyolultság

Megj: Ha U univerzális TG-k, akkor $C_M(x)$
nem függ **lényegesen** U -tól (max konstans erejéig).

A továbbiakban rögzítünk egy U univ. TG-et.

Def: Egy I^* -beli x szó esetén a szó **Kolmogorov-bonyolultsága**
 $C(x)$, ahol $C(x) = C_U(x)$.

1. Megj: $C(x) \leq |x| + k$ alkalmas, x -től független k számra.

Az invariancia tételből következik, olyan M -et választva,
mely rögtön megáll.

2. Megj: $C(\text{bin}(n)) \leq \log_2 n + c$ alkalmas c -re, ahol $\text{bin}(n)$ az n szám
bináris alakját jelöli

Kolmogorov bonyolultság

Def: Egy I^* -beli x szó **összenyomhatatlan**, ha

$$C(x) \geq |x|.$$

Tétel: Legyen $k \in \mathbb{Z}^+$ szám. Ekkor

- 1) Legfeljebb $2^{k+1} - 1$ olyan x szó van, melyre $C(x) \leq k$.
- 2) Minden n természetes számra van n hosszú, összenyomhatatlan szó.

Biz: 1) Legyen $H_k = \{x \in I^*, C(x) \leq k\}$. H_k -nak legfeljebb annyi eleme van, mint a max. k hosszú I^* -beli szorozatok száma, azaz

$$1 + 2 + 2^2 + 2^3 + \dots + 2^k = 2^{k+1} - 1$$

Kolmogorov bonyolultság

Biz: 2) Az n hosszú szavak száma 2^n .

Legyen $k = n - 1$.

Ekkor 1) szerint legfeljebb $2^{k+1} - 1 = 2^n - 1$ olyan x szó van, melyre $C(x) \leq n - 1$.

Tehát van olyan szó van, mely nincs H_{n-1} -ben, azaz

$C(x) > n - 1$, tehát $C(x) \geq |x| = n$.

Azaz x összenyomhatatlan.

1.Megj: Ha $n > 8$, akkor az n hosszú szavak több mint 99 %-ának Kolmogorov bonyolultsága nagyobb mint $n-8$.

Ui a nem ilyen sorozatok H_{n-8} -beliek, ezek száma legfeljebb $2^{n-7} - 1$, melyek aránya kevesebb mint $2^{n-7} / 2^n = 2^{-7} = 1/128 < 1/100$.

Kolmogorov bonyolultság

Tétel: A Kolmogorov bonyolultságot megadó $C: I^* \rightarrow Z^+$ függvény nem rekurzív.

Biz: Legyen $F: Z^+ \rightarrow I^*$ függvény, hogy $F(m) = x$, ahol x a lexikografikus rendezés szerinti első eleme a $\{z \in I^* \mid C(z) \geq m\}$ halmaznak.

Tegyük fel indirekten, hogy C rekurzív.

Ekkor van olyan N TG, mely kiszámítja $C(z)$ -t.

Kolmogorov bonyolultság

Ekkor legyen M az a TG, mely tetszőleges $\text{bin}(m)$ bemenet esetén:

- lexikografikus sorrendben veszi I^* elemeit,
- N -nel a kiválasztott z -re kiszámítja $C(z)$ -t,
- ha $C(z) \geq m$, megáll és kiírja z -t.

Kolmogorov bonyolultság

Ekkor $m \leq C(F(m))$ F megadása miatt.

De az invariancia tétel szerint van olyan c , hogy

$$C(F(m)) \leq C_M(F(m)) + c$$

Másrészt M definíciója szerint igaz, hogy

$$C_M(F(m)) \leq \log_2 m + 1$$

De az is igaz lesz, hogy

$$m \leq \log_2 m + c + 1$$

Ami viszont nem teljesül elég nagy m -től kezdődően.

Tehát az indirekt feltevés hamis volt, az állítás igaz.

Kolmogorov bonyolultság és a véletlen

Jelölés: I^∞ jelöli az összes végtelen 0-1 sorozatot. Egy $x \in I^\infty$ esetén x_n az első n elemből álló szót jelöli.

Def: Egy $x \in I^\infty$ sorozat **Kolmogorov-véletlen sorozat**, ha

$\lim_{n \rightarrow \infty} C(x_n) / n = 1$, azaz a sorozat kezdőszeletei „aszimptotikusan” majdnem teljesen összenyomhatatlanok.

Tétel: Ha $x \in I^\infty$ egy Kolmogorov-véletlen sorozat, akkor nincs M TG, mely n (bináris) bemenetre x_n -t állítja elő.

Biz: Indirekte, tegyük fel hogy van ilyen M TG. Ekkor $C_M(x_n) \leq \log_2 n + 1$, és az invariancia tétel szerint $C(x_n) \leq \log_2 n + 1 + c_M$:

De ekkor $\lim_{n \rightarrow \infty} C(x_n) / n \leq \lim_{n \rightarrow \infty} (\log_2 n + 1 + c_M) / n = 0$, ellentmondás!

Irodalom

11. előadás május. 7.

Rónyai-Ivanyos-Szabó: Algoritmusok, 229-238 oldalak