

Algoritmi v bioinformatiki - 2. Domača naloga

Jan Panjan

May 5, 2025

1. Dano imamo naslednje zaporedje izidov metov kovanca

$$V = CCCGCGGCGCGC$$

pri čemer C označuje, da je bil izid meta cifra, G pa da je bil izid meta grb. Za mete imamo na voljo 3 kovanice, A , B in C , veljajo naslednje verjetnosti:

Prehod:

%	A	B	C
A	40	30	30
B	30	40	30
C	30	30	40

Izpis:

%	C	G
A	75	25
B	80	20
C	20	80

Katera od možnosti je najbolj verjetna?

- (a) za vse mete smo uporabili kovanec A
- (b) za vse mete smo uporabili kovanec C
- (c) za vse mete smo uporabili kovanec B
- (d) $\Pi = AAACBCCBBCA$

Odgovor ustrezno utemeljite.

Za vse mete smo uporabili kovanec A

$(0.75)^7$	7-krat vržemo C z verjetnostjo 0.75
$(0.25)^2$	7-krat vržemo G z verjetnostjo 0.75
$(0.4)^{10}$	10-krat ne zamenjamo kovanca A z verjetnostjo 0.4

$$p(A) = (0.75)^7 \cdot (0.25)^4 \cdot (0.4)^{10} = 0.00209$$

Za vse mete smo uporabili kovanec B

$(0.8)^7$	7-krat vržemo C z verjetnostjo 0.8
$(0.2)^2$	7-krat vržemo G z verjetnostjo 0.2
$(0.4)^{10}$	10-krat ne zamenjamo kovanca B z verjetnostjo 0.4

$$p(B) = (0.8)^7 \cdot (0.2)^4 \cdot (0.4)^{10} = 0.00134$$

Za vse mete smo uporabili kovanec C

$(0.2)^7$	7-krat vržemo C z verjetnostjo 0.8
$(0.8)^2$	7-krat vržemo G z verjetnostjo 0.2
$(0.4)^{10}$	10-krat ne zamenjamo kovanca C z verjetnostjo 0.4

$$p(C) = (0.2)^7 \cdot (0.8)^4 \cdot (0.4)^{10} = 0.0000209$$

II = AAACBCCBBCA

$(0.3)^6$	6-krat ostanemo v istem kovancu (vsi kovanci imajo enake verjetnosti)
$(0.4)^4$	4-krat zamenjamo kovanec (tudi tu imajo enako verjetnosti)
$(0.75)^4$	4-krat vržemo kovanec A, vsakič vržemo cifro z verjetnostjo 0.75
$(0.8)^3$	3-krat vržemo kovanec B, vsakič vržemo cifro z verjetnostjo 0.8
$(0.8)^4$	4-krat vržemo kovanec C, vsakič vržemo grb z verjetnostjo 0.8

$$p(\text{II}) = (0.3)^6 \cdot (0.4)^4 \cdot (0.75)^4 \cdot (0.8)^3 \cdot (0.8)^4 = 0.00000124$$

Rešitev: Najbolj verjetna je možnost z največjo verjetnostjo. To je možnost (a) z verjetnostjo 0.00209.

2. Dani imamo zaporedji $s = GAGTACA$ in $t = TGATTACA$ ter vrednostno funkcijo s parametroma $\mu = 4, \sigma = 2$ in nagrado za ujemanje 2.

- (a) Z uporabo Needleman-Wunsch-evega algoritma za globalno poravnavo smo dobili naslednjo tabelo:

– slika tabele –

Dopolnite tabelo tako, da poračunate vrednosti (in ustrezne puščice) za zadnji dve vrstici.

- (b) Koliko optimalnih globalnih poravnav dobite? Izpišite vse rešitve.

3. Dano imamo naslednjo matriko izražanja:

	T_1	T_2	T_3	T_4	T_5	T_6
g_1	2	2	6	2	3	4
g_2	3	7	3	1	9	3
g_3	2	2	7	2	6	3
g_4	3	2	3	2	1	3
g_5	2	1	5	1	0	4
g_6	3	5	5	8	2	3
g_7	1	3	1	5	4	2
g_8	5	4	2	4	7	5

4. Določite gruče z uporabo metode voditeljev, če je začetna množica voditeljev enaka $X = \{g_1, g_5, g_6\}$.
5. Izračunajte drevo hierarhičnega gručenja z uporabo algoritma UPGMA.