

Bioinformatika -
Zadání 2. samostatné
práce

HMM / GE

Podmínky



Pro získání plného počtu bodů (tj. 10) je potřeba si vybrat jeden z dále uvedených úkolů a vypracovat jej

Úkoly můžete odevzdávat do 11. cvičení.

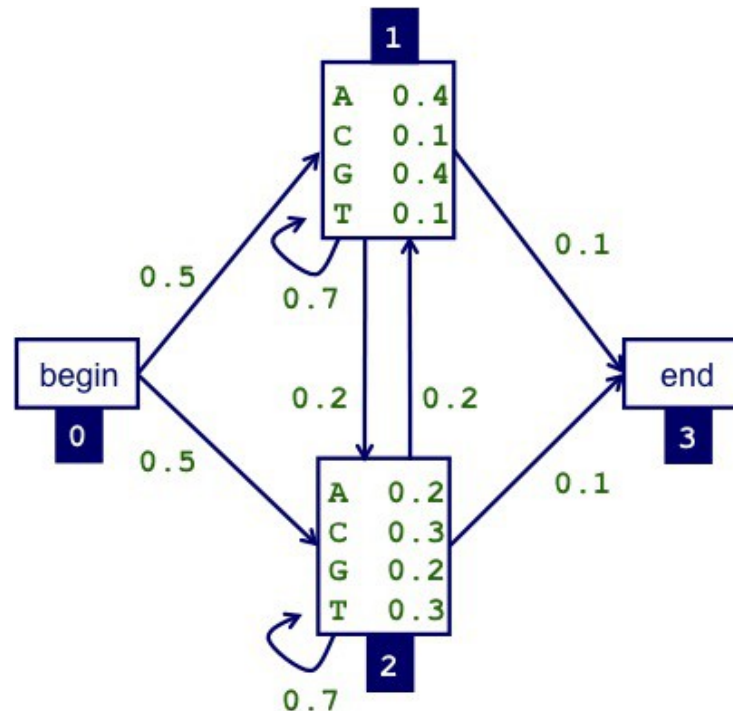
Varianta 1: HMMs

1. Naprogramujte Forward algoritmus a Viterbiho algoritmus a použijte je pro zodpovězení otázek na následujícím slajdu.
2. Není potřeba, abyste řešili uživatelský vstup. Stačí, když napíšete funkce, které převezmou na vstupu vámi zvolenou reprezentaci skrytého markovského modelu a sekvenci a na základě toho vypočítají pravděpodobnost této sekvence (Forward algoritmus), resp. nejpravděpodobnější posloupnost stavů (Viterbiho algoritmus).
3. Pozor! Je potřeba, aby vaše implementace opravdu používala dynamické programování.

Varianta 1: HMM

2. Spočítejte pravděpodobnost sekvence CGTCAG pomocí vámi implementovaného Forward algoritmu pro následující model.

3. Nalezněte nejpravděpodobnější sekvenci stavů, která vygenerovala sekvenci CGTCAG pomocí vámi implementovaného Viterbiho algoritmu.



Varianta 2: Analýza GE dat

- **Vypracujte alespoň 75% úkolů z:**
- <http://www-users.cselabs.umn.edu/classes/Spring-2012/csci3003/files/lab7/Lab7.pdf>
- **Za jeden úkol se považuje úkol nebo otázka označená malým písmenem (a), (b),..., což odpovídá v odkazovaném dokumentu podotázkám. Pokud nějaká otázka nemá podotázky, považuje se celá za jeden úkol.**

Varianta 3: Gene finding

- Náročnější (a zajímavější) úkol za 20 bodů, tj. nebudete muset dělat úkol číslo 3 (sequence assembly)
- Stačí použít hotovou implementaci Viterbiho algoritmu
- <http://www.biostat.wisc.edu/~craven/776/hw3.html>