

Bioinformatika -
Zadání 2. samostatné
práce

HMM / GE

Podmínky

Pro získání plného počtu bodů (tj. 15) je potřeba si vybrat jeden z následujících průchodů:

- HMMs (5 b.) → sestavování DNA sekvencí (10 b.)
- Gene finding (15 b.)

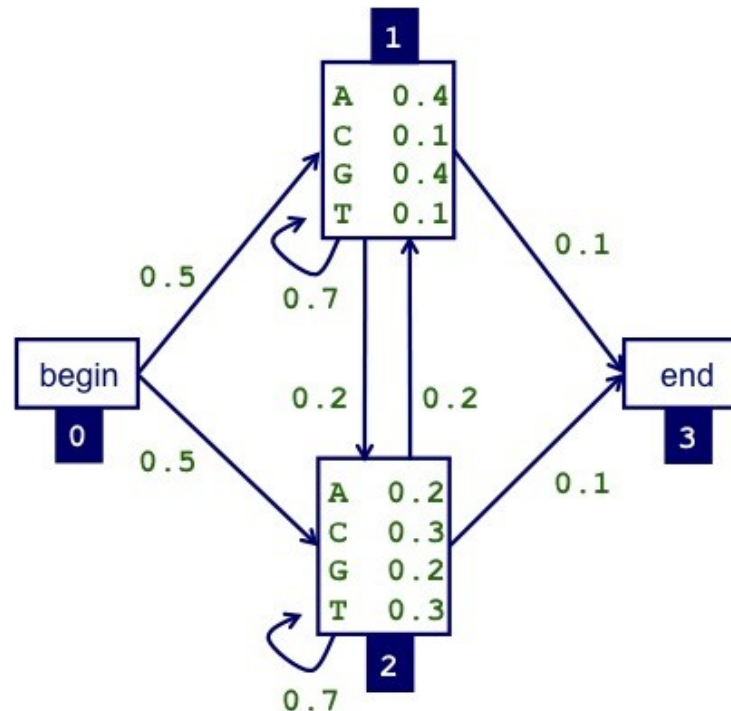
Úkoly (souhrně) můžete odevzdávat do 13. cvičení. Za každý týden navíc penalizace 3 b.

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: Gene finding

- ↩ Náročnější (a zajímavější) úkol za 15 bodů, tj. nebudete muset dělat sestavování DNA sekvencí.
- ↩ Můžete použít hotovou implementaci Viterbiho.
- ↩ <http://www.biostat.wisc.edu/~craven/776/hw3.html>