Assignment 5—Grundlagen der Bioinformatik

Anastasia Grekova 4149666 and Huajie Chen 4199962

May 27, 2018

HMM and Viterbi Algorithm

$v_l(i+1) = e_l(x_{i*1}) max_{k \in Q}(v_k(i) a_{kl})$
$v_0(0) = 1$ Probability of the Markov chain in the beginning state, that generates nothing.
$v_k(0) = 0 \; for \; k \neq 0$ Probability of the most probable Markov chain being 0 but not in the beginning state
$0 - G_P = 1 \cdot 0,5 \cdot 0,2 = 0,1$
$0 - G_Q = 1 \cdot 0,5 \cdot 0,15 = 0,075$
$\Rightarrow ptr_{p(1)} = P$
$G_P - A_P = 0,1 \cdot 0,7 \cdot 0,4 = 0,028$
$G_Q - A_P = 0,075 \cdot 0,2 \cdot 0,4 = 0,006$
$G_P - A_Q = 0,1 \cdot 0,3 \cdot 0,1 = 0,003$
$G_Q - A_Q = 0,075 \cdot 0,8 \cdot 0,1 = 0,006$
$\Rightarrow ptr_{p(2)} = P$
$A_P - T_P = 0,028 \cdot 0,7 \cdot 0,1 = 0,00196$
$A_Q - T_P = 0,006 \cdot 0,2 \cdot 0,1 = 0,00012$
$A_P - T_Q = 0,028 \cdot 0,3 \cdot 0,35 = 0,00294$
$A_Q - T_Q = 0,006 \cdot 0,8 \cdot 0,35 = 0,00168$
$\Rightarrow ptr_{q(3)} = P$
$T_p - A_P = 0,00196 \cdot 0,7 \cdot 0,4 = 0,0005488$
$T_Q - A_P = 0,00294 \cdot 0,2 \cdot 0,4 = 0,0002352$
$T_P - A_Q = 0,00196 \cdot 0,3 \cdot 0,1 = 0,000058$
$T_Q - A_Q = 0,00294 \cdot 0,8 \cdot 0,1 = 0,0002352$
$\Rightarrow ptr_{p(4)} = P$

Therefore, the most probable approach should be :

$$GATA$$
$$PPPP$$

v		G	A	T	A
0	1	0	0	0	0
A_p	0	0	0,028	0	0,0005488
C_p	0	0	0	0	0
G_p	0	0,1	0	0	0
T_p	0	0	0	0,00196	0
A_q	0	0	0,006	0	0,0002352
C_q	0	0	0	0	0
G_q	0	0,075	0	0	0
T_q	0	0	0	0,00294	0