Pengfei Li

HMM for DNA Sequence

for the followings, "/M" is omitted

$$P(x|M) = P(CGTCAG|M) = \sum_{i=1}^{2} \alpha_6(i) = \sum_{i=1}^{2} P(O_1O_2\cdots O_6 \wedge (q_6 = S_i))$$

as we know

$$\alpha_t(i) = P(O_1O_2\cdots O_t \wedge (q_t = S_i)|M)$$

$$\alpha_1(i) = b_i(O_1)\pi_i$$

$$\alpha_{t+1}(j) = \sum_i a_{ij} b_j(O_{t+1})\alpha_t(i)$$

Thus

$t=1$, $i=1,2$

$\alpha_1(1) = b_1(O_1)\pi_1 = 0.1\times 0.5 = 0.05$ $\alpha_1(2) = b_2(O_1)\pi_2 = 0.4\times 0.5 = 0.2$

$t=2$, $i=1,2$

$\alpha_2(1) = \sum_i a_{i1} b_1(O_2)\alpha_1(i) = 0.8\times 0.4\times 0.05 + 0.2\times 0.4\times 0.2 = 0.032$

$\alpha_2(2) = \sum_i a_{i1} b_2(O_2)\alpha_1(i) = 0.2\times 0.1\times 0.05 + 0.8\times 0.1\times 0.2 = 0.017$

$t=3$, $i=1,2$

$\alpha_3(1) = \sum_i a_{i1} b_1(O_3)\alpha_2(i) = 0.8\times 0.1\times 0.032 + 0.2\times 0.1\times 0.017 = 0.0029$

$\alpha_3(2) = \sum_i a_{i2} b_2(O_3)\alpha_2(i) = 0.2\times 0.4\times 0.032 + 0.8\times 0.4\times 0.017 = 0.008$

Inductively

$t=4$, $i=1,2$

$\alpha_4(1) = 0.8\times 0.1\times 0.0029 + 0.2\times 0.1\times 0.008 = 0.000392$

$\alpha_4(2) = 0.2\times 0.4\times 0.0029 + 0.8\times 0.4\times 0.008 = 0.002792$

$t=5$, $i=1,2$

$\alpha_5(1) = 0.8\times 0.4\times 0.000392 + 0.2\times 0.4\times 0.002792 = 0.0003488$

$\alpha_5(2) = 0.2\times 0.1\times 0.000392 + 0.8\times 0.1\times 0.002792 = 0.0002312$

$t=6$, $i=1,2$

$\alpha_6(1) = 0.8\times 0.4\times 0.0003488 + 0.2\times 0.4\times 0.0002312 = 0.00013012$

$\alpha_6(2) = 0.2\times 0.1\times 0.0003488 + 0.8\times 0.1\times 0.0002312 = 2.5472\times 10^{-5}$

To sum up, $P(x|M) = \alpha_6(1) + \alpha_6(2) = 0.00015558$