

情報基礎実験 02

DP

05-175511 中林亮

1. 初期値

M_{ij} は $i-1, j-1$ 文字目までを使うアライメントで、最後がマッチ・ミスマッチで終わっているものの点数の最大値。 I_{ij}^x, I_{ij}^y は $i-1, j-1$ 文字目までを使うアライメントで、最後がギャップで終わっているものの点数の最大値。

$M_{00} = 0$ とするとき、

M_{11}

空文字と～のアライメント

$$M_{i,0} = -d - (i-1)e \text{ if } i > 0$$

$$M_{0,j} = -d - (j-1)e \text{ if } j > 0$$

$$I_{i,0}^x = -\infty, I_{0,j}^x = -\infty, I_{i,0}^y = -\infty, I_{0,j}^y = -\infty$$

2. 対称性

更新式は

$$M_{i+1,j+1} = \max(M_{i,j}, I_{i,j}^x, I_{i,j}^y) + c_{s_i,s_j} I_{i+1,j}^x = \max(M_{i,j}-d, I_{i,j}^x-e, I_{i,j}^y-d) I_{i,j+1}^y = \max(M_{i,j}-d, I_{i,j}^y-e)$$

非対称になっている項は、 I^x の 3 項目に対応する I^y の、 $I_{i,j}^x - d$ となるはずの項。つまり、x がギャップになる直後の y のギャップは認めるが、y がギャップの直後の x のギャップは認めていない。例えば以下の例では case1 を認める一方 case2 を認めない:

case1:

x: - C

y: A -

case2:

x: C -

y: - A

アライメントは $(x_i, y_j), (x_i, -), (-, y_j)$ の組の集合だと考えれば、上のケースの後に続く組みは上記のマッチ or ミスマッチ・y ギャップ・x ギャップの 3 通り。

(1) マッチ or ミスマッチの時、case1 と 2 の点数は同じ。

(2) y ギャップの時

3. 実装

NWG.py の通り

4. 確率計算

1: AC

2: AG