Алгоритмы в биоинформатике: Домашнее задание № 2, теоретическая часть

Выполнила: Костина Ю.М., НИУ ВШЭ, группа МВБ 221

15 февраля 2023 г.

1. Количество выравниваний

   Выведите рекуррентную формулу количества всех возможных выравниваний последовательностей длины n и m пользуясь разбиением всех выравниваний на непересекающиеся блоки. (1.5 балл) Получите точную формулу, основываясь на начальные условия и рекуррентную формулу. (1.5 балл) Воспользуйтесь приближением Стирлинга чтобы получить приближенную формулу количества выравниваний. (1)

   Решение:

   (a) Пусть количество всех возможных выравниваний последовательностей $a_1 \dots a_n$ и $b_1 \dots b_m$ обозначается как $f(n, m)$.

   Пусть $|F| = f(n, m)$ - множество всех возможных выравниваний. Тогда:

   i. $|F_1| = f(n-1, m-1)$ - все выравнивания для символов $a_i = b_i$
   ii. $|F_2| = f(n, m-1)$- все выравнивания где $a_i = \emptyset$
   iii. $|F_3| = f(n-1, m)$- все выравнивания где $b_i = \emptyset$
   iv. $|F_4| = |F_2 \cap F_3| = f(n-1, m-1)$ - пересечение V_2 и V_3, так как выравнивание вида $a_i = \emptyset \cap b_i = \emptyset$ эквивалентны

   Отсюда:

   $$f(n,m) = f(n-1,m-1)+f(n,m-1)+f(n-1,m)-f(n-1,m-1) = f(n,m-1)+f(n-1,m)$$

   (b) Если в одной последовательности 1 элемент и в другой - нет элементов, то справедливо $f(1, 0) = C^0_{1+0} = 1$, $f(0, 1) = C^1_{1+0} = 1$ - наши начальные условия. По предположению индукции условия выполняются для $f(n, m-1)$ и $f(n-1, m)$:

   $$f(n, m-1) = f(n-1, m-1) + f(n, m-2) = C^{n-1}_{n+m-2} + C^n_{n+m-2} = C^n_{n+m-1} = \frac{(n+m-1)!}{n!(m-1)!}$$

   $$f(n-1, m) = f(n-2, m) + f(n-1, m-1) = C^{n-1}_{n+m-2} + C^{n-1}_{n+m-2} = C^{n-1}_{n+m-1} = \frac{(n+m-1)!}{(n-1)!m!}$$

   тогда из правил комбинаторики выполняется индукционный переход:

   $$f(n, m) = f(n, m-1) + f(n-1, m) = C^n_{n+m-1} + C^{n-1}_{n-1+m} = C^n_{n+m} = \frac{(n+m)!}{n!m!}$$

   что и требовалось доказать