

به نام خداوند بخشنده و مهربان



دانشگاه صنعتی امیرکبیر  
( پلی تکنیک تهران )

## درس مقدماتی بر بیوانفورماتیک

گزارش بخش برنامه‌نویسی تمرین دوم

استاد درس: دکتر زینعلی

نام دانشجو:

روزبه قاسمی ۹۵۳۱۴۲۴

فروردین ۱۳۹۹

برای پیاده سازی الگوریتم **star** باید سه مرحله اصلی آنرا پیاده سازی کرد تا نهایت وقتی چند توالی به برنامه می‌دهیم، مطابق خواسته مسئله و بر اساس امتیازات **match**، **mismatch**، **gap open** و **gap** **extension** هم‌ترازی‌ها بدست می‌آیند. مراحل آن مطابق اسلاید زیر است:



## الگوریتم Star

1. یک توالی را به عنوان توالی مرکزی انتخاب کنید.
  2. هر توالی را با توالی مرکزی هم‌تراز کنید.
  3. هم‌ترازی‌ها را ادغام کنید.
- با قانون "once a gap, always a gap"



## مرحله اول الگوریتم Star

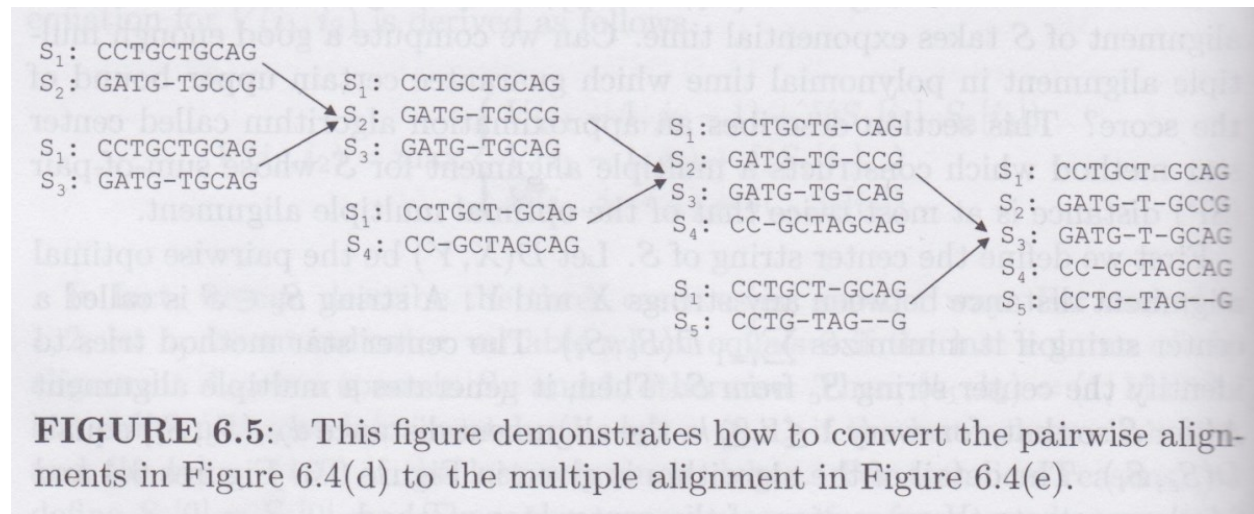
برای پیدا کردن توالی مرکزی باید بر اساس برنامه تمرین قبلی هر زوج رشته را به برنامه قبلی بدهیم تا امتیاز بدست آمد، سپس **ماکزیمم جمع امتیازهای هر سطر، برابر با توالی منتخب ماست.** بر اساس تست کیس داده شده جدول زیر بدست می‌آید (طبق خواسته گفته شده در کانال بر همان اساس امتیازات هر مرحله به دست می‌آید و امتیازات **match** و ... بر اساس امتیازات گفته شده در تمرین قبل است):

	S1	S2	S3	S4	Max
S1		-11	-13	-15	-39
S2	-11		-14	-11	-36
S3	-13	-14		-12	-39
S4	-15	-11	-12		-38

بنابراین توالی دوم به عنوان توالی مرکزی انتخاب می‌شود.

## مرحله دوم الگوریتم Star

سپس در این مرحله باید دوباره دو مطابق الگوریتم زیر باهم ادغام شود، بر این اساس که همانند مثالی که جدول آن را در مرحله اول به دست آوردیم، با هم همتراز می‌کنیم:



**FIGURE 6.5:** This figure demonstrates how to convert the pairwise alignments in Figure 6.4(d) to the multiple alignment in Figure 6.4(e).

## مرحله سوم الگوریتم Star

سپس آن را باهم ادغام می‌کنیم:

```
if sum_list[1] == max_val:
    # print("The Sc is second sequence: ")

    res_seq_0_2 = Affine_gap(GA_second_seq_1_0, GA_second_seq_1_2)
    GA_second_seq_1_0, GA_second_seq_1_2, num_seq_1_0 = res_seq_0_2.alingment()
    res_seq_0_2 = Affine_gap(GA_second_seq_1_0, GA_second_seq_1_3)
    GA_second_seq_1_0, GA_second_seq_1_3, num_seq_1_0 = res_seq_0_2.alingment()
    res_seq_0_2 = Affine_gap(GA_second_seq_1_2, GA_second_seq_1_3)
    GA_second_seq_1_2, GA_second_seq_1_3, num_seq_1_0 = res_seq_0_2.alingment()
    res_seq_0_2 = Affine_gap(GA_first_seq_1_0, GA_second_seq_1_3)
    GA_first_seq_1_0, GA_second_seq_0_3, num_seq_1_0 = res_seq_0_2.alingment()

    print(GA_second_seq_1_0)
    print(GA_first_seq_1_0)
    print(GA_second_seq_1_2)
    print(GA_second_seq_1_3)
```

## تست برنامه

خروجی تست برنامه آن در نهایت به شکل زیر است:

```
Run: HW2_9531424_Q4 x
D:\Anaconda\python.exe "C:/Users/USER/Desktop/Term 8/BioInfo/Homeworks/HW2/HW2_9531424_Q4.py"
CTCTGGAAGT
ATCGGAT
AACCTAGG
ACTGACTG
CTCTGGAAGT
ATCGG---AT
AAC-CCTAGG
--ACTGACTG
Process finished with exit code 0
```

## چالش های پیاده سازی

بر اساس خروجی هایی که برای تست های مطرح شده در کوئرا قرار دادید، خروجی من با خروجی جواب یکی نیست و نمره ۰ به من تعلق گرفته است. این درحالی است که مشکل نه از پیاده سازی من است و نه از خروجی شما بلکه علت آن باز می گردد به قسمت **Alignment** کردن تمرین قبلی چون که اگر توجه نمایید، برای **Traceback** کردن ما مطابق اسلاید زیر عمل می کنیم:



همترازی سراسری با استفاده از affine gap

$$M(i, j) = \max \begin{cases} M(i-1, j-1) + s(x[i], y[j]) \\ I_x(i-1, j-1) + s(x[i], y[j]) \\ I_y(i-1, j-1) + s(x[i], y[j]) \end{cases}$$

$$I_x(i, j) = \max \begin{cases} M(i-1, j) - d \\ I_x(i-1, j) - e \end{cases}$$

$$I_y(i, j) = \max \begin{cases} M(i, j-1) - d \\ I_y(i, j-1) - e \end{cases}$$

حال اگر دقت کنید متوجه می‌شوید که به طور مثال برای ماتریس  $M$  وقتی که ما در زمان بدست آوردن ماکزیمم، بیشتر از یکی از آنها درست شود، باعث می‌شود که شرط `if` یا `elseif` شود دیگر بقیه چک نمی‌شود و مشکل از همینجاست!

تنها راه حلی که به ذهن من می‌رسید این بود که هر  $12$  حالت ( $3$  حالت برای  $M$ ،  $2$  حالت برای  $ix$  و  $2$  حالت برای  $ly$  پس در نهایت  $2*2*3$  حالت)، باید بررسی شود سپس انتخاب کنیم که کدام هم‌ترازسازی مطابق میل ماست. اما از آنجایی که این روش بسیار روش `Inefficient` ای به حساب می‌آید زیرا بسیار حل مسئله آن در این حالت زمان‌بر خواهد بود و راه دیگری به ذهنم نمی‌رسید برای همین هم‌ترازسازی هایی که من بدست آوردم، درست است و امتیاز آن نیز به درستی به دست آمده اما یکی از حالت هایی است که شما آنرا در نظر نگرفته اید. بنابراین پیاده سازی من کاملاً درست و منطقی است اما از لحاظ کوئرا نمره ای به آن تعلق نمی‌گیرد!

پایان