REPORT



**알고리즘**

**Algorithm**

Assignment 3

황호영 교수님

컴퓨터정보공학부

2017202037 오민혁

2020/12/02 (수)

**< Problem Set #3 >**

이번 과제는 a longest common subsequence(LCS) 에 관한 것이다.

이는 두 개의 길이가 n 인 수열 이 있다고 가정한다.

1. (a) 우선 다양한 문자열의 길이에 대한 LCS를 실행 하기 위해 LENGTH List를 만들었다.

이 LENGTH LIST는 5, 10, 15, 20, 25, 30, 35, 40 으로 구성되어 있다.

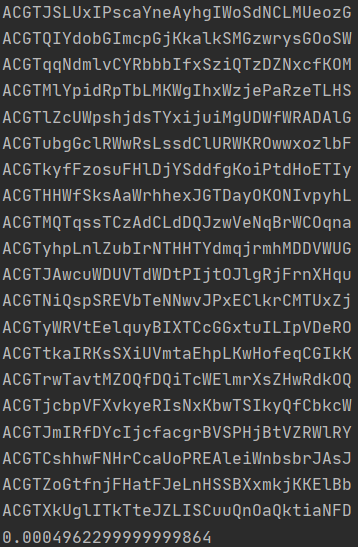
그 다음 이 각각의 LENGTH 길이를 가진 두 문자열을 생성하여, 이 두 문자열의 LCS 값을 구하는데 걸리는 시간의 평균 값을 측정하기 위해 10번의 랜덤 문자열 생성과, 이에 대한 LCS가 진행되도록 하였다.

본 과제를 수행하기 위해 a longest common subseqence 알고리즘을 코드로 작성하였다. 이는 bottom-up 방식으로 dynamic-programming 로 구현하였는데, bottom-up을 위해 가로 length2+1, 세로 length1+1 길이의 2차원 리스트를 선언하였다.

그 다음 2중 반복문을 통해 두 문자열을 비교하는데, 만약 가장 최근에 추가 된 문자가 같다면 추가 하기 이전의 단계의 LCS 길이 즉 LCS[i-1][j-1] 의 길이에 +1을 더한 값을 LCS[i][j]에 기록해준다.

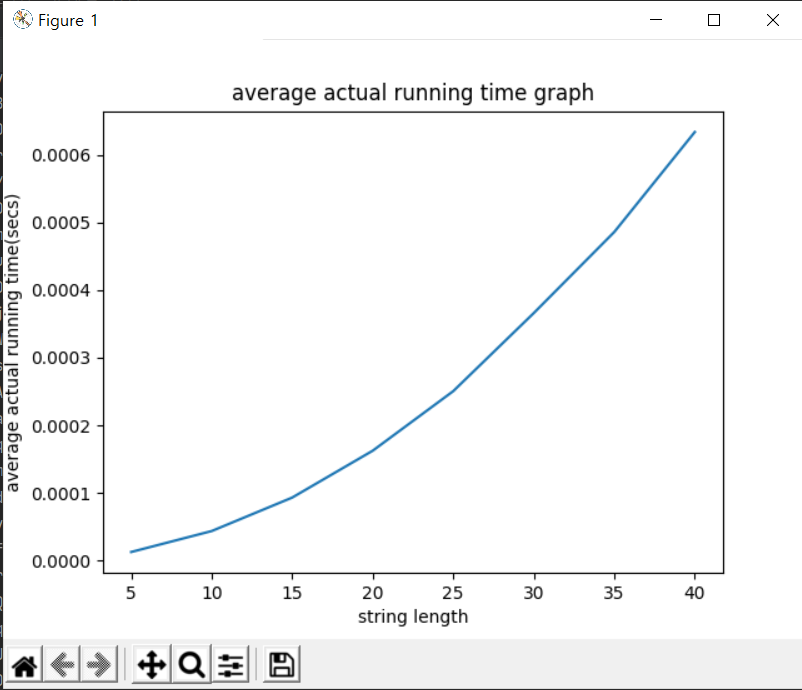
만약 같지 않다면, LCS[i-1][j] 와 LCS[i][j-1] 두 값 중에 더 큰 값을 LCS[i][j]에 기록해주는 방식으로 구현하였다.

(b)

좌측의 스크린샷은 LENGTH가 35일 때, 두 개의 무작위 A,C,G,T 를 sub sequence로 가지고 있는 Random string 을 생성하고, lcs를 구하는 작업을 10번 반복하는데서 나온 Input string 들이다. 맨 밑의 0.0004962는 이 문자열들에 대한 LCS를 구하는데 소요되는 시간의 평균 시간이다.

(c) 수행 시간은 LCS 2차원 리스트를 만드는 부분부터 LCS를 구하기 위한 dynamic programming을 수행하는 부분까지를 측정한 것으로 하였다. 이외의 랜덤 한 문자열을 생성하거나 하는 부분은 실제 LCS 알고리즘의 수행 시간이 아니라고 판단하였기 때문에 위와 같이 수행 시간을 측정하였다.

|  |  |
| --- | --- |
| **n** | **평균 수행 시간(sec)** |
| 5 | 0.00001286 |
| 10 | 0.00004383 |
| 15 | 0.00009331 |
| 20 | 0.0001625 |
| 25 | 0.0002505 |
| 30 | 0.0003659 |
| 35 | 0.0004857 |
| 40 | 0.0006334 |



**< 컴퓨터 사양 >**

CPU : Intel® Core™ i7-8750H CPU @ 2.20GHz 2.20 GHz

RAM: 16.0GB

OS: Windows 10 HOME 64bit

[참고문헌]

<https://mingrammer.com/underscore-in-python/> / 파이썬 언더스코어(\_)에 대하여

[https://suri78.tistory.com/11](https://suri78.tistory.com/11%20/)  [백준알고리즘] 9251번: LCS – Python

<https://dojang.io/mod/page/view.php?id=2293> / Python 반복문으로 리스트 만들기

<https://hongku.tistory.com/297> / python 랜덤 문자열 생성

<https://bskyvision.com/760> / python 코드 수행시간 측정

<https://wikidocs.net/92080> / matplot 사용법