
Μέγιστη κοινή υποακολουθία (LCS=Longest Common Subsequence)

Παναγιώτα Βίννη Α.Μ. 1873
29 Ιανουαρίου 2022

Περίληψη

Σε αυτήν την τεχνική αναφορά εξετάζεται η εύρεση της μέγιστης κοινής υποακολουθίας DNA με τον δυναμικό προγραμματισμό, χρησιμοποιώντας τους αλγόριθμους Brute Force και LCS (Last Common SubSequence).

1 Εισαγωγή

Το πρόβλημα της Longest Common Subsequence είναι ένα πρόβλημα στο οποίο δίνονται οι ακολουθίες DNA (στην περίπτωση μας βγαιίνουν τυχαία μέσω της random) και ζητείται να βρεθεί η max κοινή υποακολουθία από όλες τις ακολουθίες και το μέγεθός της.

Για τη σχεδίαση του χρησιμοποιήθηκαν 2 αλγόριθμοι. Πρώτος είναι ο άπλοϊκός αλγόριθμος Brute Force (ωμής δύναμης), ο οποίος δημιουργεί όλες τις υποακολουθίες της μίας ακολουθίας DNA και ελέγχει αν υπάρχει στην άλλη. Και τέλος, είναι ο αλγόριθμος LCS (δυναμικού προγραμματισμού), όπου κάνει ακριβώς την ίδια διαδικασία απλά φτιάχνει επιπλέον και έναν πίνακα απομνημόνευσης αποτελεσμάτων με ακέραιους αριθμούς, οι οποίοι αυξάνονται κάθε φορά που είναι κοινό το γράμμα στις ακολουθίες.

2 Αποτελέσματα

Για την συγγραφή της εργασίας χρησιμοποιήθηκε η Python 3.9.2 και το Visual Studio Code. Τα χαρακτηριστικά του υπολογιστή είναι Intel Core i7-1065G7 (1.30GHz - 1.50 GHz), 8 GB DDR3 (2667 MHz), λειτουργικό σύστημα Windows 10 Home x64. Τέλος, οδηγίες για την εκτέλεση του κώδικα: `python main.py` και για την εκτέλεση των tests : `python tests.py`

Subseurnces Length	10
Subsequences Characters	20
Lenght of LCS	14
LCS	GGGACATGGTGCC

Πίνακας 2.1: Πίνακας αποτελεσμάτων

Σχήμα 2.1: Αποτελέσματα εκτέλεσης κώδικα.

```
C:\Users\penyv\OneDrive\Υπολογιστής\DIT\Εξάμηνο 5ο\Αλγορίθμοι & Πολυπλοκότητα\erg3\assignment_3>pytho
['ATCCGATTCCCTGTAAGGGG', 'AGAGGACTTCCGCGCCGAGT', 'TAATCACCGCATCACGGTG', 'GTAAGCCATTCCTGCCCC', 'ACT
GCC', 'GGCGACATGGATCGACTTCG']
Brute forse method:
[('AGAGGACTTCCGCGCCGAGT', 'ACTCGACACTACCCGCCGGC', 'AGAACTCCCGCCGG', 14), ('AGAGGACTTCCGCGCCGAGT', 'AG
14)]
LCS method:
[('AGAGGACTTCCGCGCCGAGT', 'ACTCGACACTACCCGCCGGC', 'AGAACTCCCGCCGG', 14), ('AGAGGACTTCCGCGCCGAGT', 'AG
14)]
Length of longest common sub-sequence:
14
Longest common sub-sequence
GGGACATGGTGCC
```

3 Συμπεράσματα

Η εύρεση μέγιστης κοινής υπο-ακολουθίας είναι ένα ενδιαφέρον πρόβλημα που υλοποιείται κάνοντας χρήση του δυναμικού προγραμματισμού, και συγκεκριμένα των αλγορίθμων Brute Forse ή LCS.