CISC-471 FALL 2019

HOMEWORK 4

Please work on these problems and be prepared to share your solutions with classmates in class on Thursday Oct. 4. Assignments will **not** be collected for grading.

Programming

Write a program in the language of your choosing (I recommend Python) and verify that it works on the sample data (using the on-line Rosalind platform). For each problem be prepared to tell us why you think your algorithm is correct (whether you program worked on the sample data or not). Also provide an estimate of the time and space complexity of your algorithm.

Greedy Motif Finding:

A greedy heuristic for finding motifs, GREEDYMOTIFSEARCH, is described in section 5.5 of the text. Implement this algorithm and try it on the following data:

AAATTGACGCAT
GACGACCACGTT
CGTCAGCGCCTG
GCTGAGCACCGG
AGTACGGGACAG

and find the best 3-mer motif.

Problem 5.18: Design an input for GREEDYMOTIFSEARCH algorithm that causes the algorithm to output an incorrect result. That is, create a sample that has a strong pattern that is missed because of the greedy nature of the algorithm. If optimalScore is the score of the strongest motif in the sample and greedyScore is the score returned by GREEDYMOTIFSEARCH, how large can optimalScore/greedyScore be?

Problems

Not from the text: The BreakpointReversalSort algorithm is described in our text book as follows:

BreakpointReversalSort(π)

while b(π) > 0

Among all reversals, choose reversal ρ minimizing $b(\pi \cdot \rho)$

$\pi \to \pi \cdot \rho$

output π

return

Consider the following input π with sentinel values 0 and 7.
$\boxed{0}$ 4 5 6 1 2 3 $\boxed{7}$
There are 3 breakpoints in this input. Show that there is an infinite sequence of reversals such that the number of breakpoints remains static at 3.

Problem 5.4: (modified) There are two algorithms described in our text book that solve the sorting by reversals problem, with the step
Among all reversals, choose reversal ρ minimizing $b(\pi \cdot \rho)$
This step is ambiguous because it does not say what to do if two or more reversals yield the same number of breakpoints. Consider the following algorithm where this ambiguity is removed.
BPREVSORT(π)

while b(π) > 0

if π has a decreasing strip

Choose the decreasing strip containing the smallest element k.

Since element $k-1$ must be in an increasing strip, reverse the substring that results in k and $k-1$ becoming adjacent.

else (no decreasing strip)

Choose the increasing strip containing the smallest element $k \geq 1$ (this avoids the sentinel) and reverse it to create a decreasing strip.

output π

return

Perform the BPREVSORT algorithm with $\pi = 3\ 4\ 6\ 5\ 8\ 1\ 7\ 2$ and show all intermediate permutations. Since BPREVSORT is an approximation algorithm, there may be a sequence of reversals that is shorter than the one found by BPREVSORT. Could you find such a sequence of reversals? Do you know if it is the shortest possible sequence of reversals? Compare and contrast BPREVSORT with the algorithms in the text BREAKPOINTREVERSALSORT and IMPROVEDBREAKPOINTREVERSALSORT.

Problem 5.5: Find a permutation with no decreasing strips for which there exists a reversal that reduces the number of breakpoints.