Final Project Proposal

Alice Huang (aliceh4)

Type of Project

I plan on implementing Nussinov's algorithm, an algorithm we discussed in class. I will have all my work available on GitHub (I have not created a repository for it yet).

Brief Motivation

In HW 3, we were asked to write out values found based on Nussinov's algorithm for a given sequence. Doing it by hand while referencing the recursive definition of the algorithm was pretty straightforward, so I think it would be interesting to try to implement it myself. The given recursive structure of the algorithm also leads to a dynamic programming solution, something I want to get more practice in doing.

The prediction of secondary RNA structures is also just an interesting and important problem in bioinformatics, so it would be fulfilling to see how it is done by working on this algorithm.

Dataset

I will be using data to test my implementation of the algorithm. I will get the data from the RCSB Protein Data Bank – a subset of RNA sequences along with their dot-parentheses notations will be selected to test my algorithm code.

Planned Method and Testing

I plan on following a similar structure to the coding assignments given for this class. I will create a Google Colab notebook with the necessary code/imports and write out test cases similar to the ones given to us in our homework. I will use the dataset mentioned above and loop through the sequences to test the algorithm score vs. the actual score (# of parentheses). I will only consider my algorithm "finished" if I can loop through the dataset and not find any case where the algorithm scores != the # of parentheses. I will also check the matrix values, not just the final score, for the example that we completed in our homework.

Timeline

Nov 10, 2022 : Submit proposal

Nov 14, 2022 : Write out pseudocode for the algorithm (DP)

Nov 18, 2022 : Figure out how to import/prepare the dataset for implementation

Dec 2, 2022 : Write out code to fill out the DP table and test using known solutions

Dec 9, 2022 : Write out the rest of the code and write some test cases

Dec 14, 2022 : Work on any finishing touches in the code and start writing the final report

Dec 18, 2022 : Finish the final report and submit