CSE 4065

Computational Genomics

-Project 2-

Dynamic Pairwise Sequence Alignment

Report

Melisa DÖNMEZ - 150116030

Kevser İLDEŞ - 150116048

- **Information About Implementation**

In this project, we implemented the dynamic sequence alignment algorithm with gap extensions.

Firstly, we initialized match, mismatch, gap opening and extension scores on top of the code. Then in main method, we took the file name from the user and process it to take 2 sequences in array. After that, we instantiated the SequenceAlignment class implemented for alignment operation with those parameters and called the runAlignment method of this class.

In this SequenceAlignment class, firstly we assigned class variables through parameters given and create 3 m x n matrices, where m and n values are the length of the sequences, for middle, upper and lower matrices.

In runAlignment method of this class which runs the main operations, firstly it initializes the matrices by calling the initMatrices method of the class, in which first rows and columns of all 3 matrices initialized accordingly, then in a nested for loop, it goes through sequences and fill the matrices and for each node assignment, it checks the node it came from and fill the backtracking matrix with index of it to use later in alignment construction. Finally, after all matrices are filled, it calls the constructAlignment method of the class and prints the resulting alignment sequences and its score.

- **Results**

Here are results for sample test cases:

1. <u>test1.seq:</u>

```
Enter the name of the input File: test1.seq

********** Result **********

Sequence 1:  CGG--GTAGTTAACCCT-ACA-GCATAGAGTCGCGAGATAAAGTGCAGGA-GTCTTTCGCGGCAGATTCGTACCTCA---ACCACGTGCTACTT---TCTGGCATCACGAATCTGCCGCATAGGTCCGTGAGT-CCATATGA
Sequence 2:  AGGAAGTAGTTAGCC-TAACAGGCATAGAGTCGCGACATAT-GTGAAG-ATGTCATTCG--GTA--TTCAAACCTCATGCATCAT-TGC--CTTGAGTC--GC-TC------CTGGAGCATAG-TCCCTGAGTGCCATATGA
Score     :  153.0
```

2. test2.seq:

```
Enter the name of the input File: test2.seq

********** Result **********

Sequence 1:  -C-GAGACCGA-C-GAAGAGGTT--TGGCCCCAA-CCAGGTTCC-CTGATCACGTAACTT----ACCGGCCAAAAGGACTG--GCCTT-ACTAAGGC--CTT----TGTCTACTGC-G-G--GT-C--CGG-G-GGC-CGTT--GGT-T-TC-GGCAGAACACTTC
Sequence 2:  CCTG-GACCGAGCTTAA-A--TTGCTAGC---AATACAGATGCCGCT--TC------CTTGGGGAG-GGTGTGTAGGA-TGTAGG-TTAACGAATGCAAGTTCCGGGGTAT-C-GCAGAGTCGTGCTACGGCGTGGCAC-TTAGGGTCTCTCGGGAAAAAGAGTAG
Score     :  103.0
```

3. test3.seq:

```
Enter the name of the input File: test3.seq

********** Result **********

Sequence 1:  GTGT-GGTTGCTTGCATCACTCCGTGTACATGTGACAACCGAACGAGTTGATCCAGCTTTGTTAAGTCAGCTTCGAATG-CGGTAGCTCTCAA-ATATGAT-ATG-AC-TCTTGGGGTAGAT--GCTGG--GGACCTATTGCGC-CC----A-AAGCGA-TATTCGGGGCA-
CCGGTTTAGGGTACCCTATCAAGGCGAT-AC-TTCGATGCAGTGTG-ATGCCGG-A-G-GTGTCG-GTCAGCATGTGAGA--GCT
Sequence 2:  G-GTAGGTT-CGTGAAGCACTCC-TGGAC-TCTGACCAC--AAC-A--TAATCAAGC---GC--AG--AG-TT-GAATGACCAAAGCGCTCAAGCT-T--TCACGAACGTCTC-----ATATCCGC-GGTCGTACAAAC-GCGCTCTTAAAACAATCGTCTAT-C----
CATCCGG----GGATACCA-AT---GG-G-TGACATT--A---ACTGGGCAGGCCGGCACGCG-GACGCGACAG-AT-TGAAATGGAT
Score     :  196.0
```

4. test4.seq:

```
Enter the name of the input File: test4.seq

********** Result **********

Sequence 1:  CG-GGGAAAGA-CGG--A-ATGCATCG-ACCATCGGACAAT-GC-CTCACTGGAAGCGCTG--CTGATTTTTGCGCA-ACGAGCCTCGGA-CCTCCCG-C-TCAAA-CTT-ACGAA--A-ATGACTCCACC-----AG-CACTGAA-CCAACTGGCCTCCAG-TGA-AGAT---
AGTAT-CTACAA--ATC-GTTTGCCGGGAGAAACATTT-GT-A-TGGTAGTCAGCGTTCTGCACGTCACGTAATCGT--TCACTAGTT-GTGG-GGTA---TACCAGGTC---GTAATGAG-----A-TG--TTAGT--A-TGAATCGTTTATAAC-GCTTGTCATAGGAG-TTC---C-
GAATAATCGTCA-CTAGGTC-AAT-G---GCCCCCTAC-TTGTA--ATAACTAC-GTCTGATTGAG-AAA---CAGATCGTTAGTCA-TCGGTT-AATAGTCCCGGAAGATAACGGGTTCTTGCGTTTTTGCGAATACTACTAT-CTCATGGCGATGG--A-G---C-GT---T-T---GGTCCCAT-
CC---AG-CC-GCGC-GAGTATCACTTGTTCGCCTGC-AC--CT-GTCCGACCTT-TC--GA--TGGG---GAGCTCCTTTCATTGG--TTGTAG-G-TACTAAGGGTGAGCAATGTCACGTGACCCAA-GGA-GCCGTGTGTAATTTCCACTTGCTCAGAAAAG--CCTC-GAC---A-AT--CTG-GG-
A-CCGACACCTGAGTG-AT--GGCTTACA-GACCAGGGGAG-GG-GGGCAGGTTCCCTGGCCACAGAATGGCACGCCCTGAGGAGGCACCGGCCCAA-CGTCG-CTGG
Sequence 2:  AGTGAGAA-GACCGGATATATGCAACGAACAATCGCAAAATAGCTC-CA-TGTACGC-CTGTCCCG-TAATACCGCATGCGAGCCTCGAAGCC-CCCGACGTCAAACCTCAAC-AAGTATAGGGTTCCACCGTATGAGCCACAGAATTCAGTT--CCTCC-GCTGATAG-TCCCGGT-
TCCTCCAATCAGCAGATT---CTG-----AC-TTAGGTCACTGGAAGAAA-C---CTG---G---CGTATT-GTGATGAA-ATTTCGTGGTGG-ACGCTACC--GTCTAGGTA--GAGCCACTAGTGCATT-GTACACTG--TCTTTT-TC-CGGCTA-TAATCGGAGTTTCAGGCTGACT--T-
GCCACCCAGGTCAAATAGTATGTC----ACGGTGTATGATCTCT-CGGGCTGAT--AGCAAGGGGCGG--CGTCGGGCACTCC-TTGAATTGACC------ATTT--GG--CGTGCCTTGTTGC-----CTCCT-TCCTCAAG-C-AGGGTTAGGCTTCAGTAAGTGTGGGGGTGGC-
TGCCGAAAGACGGGCTCGGGGT-------GTT-GCC--CAACGTCTGGTC-GAC-TTCCCATAATCTGGGCACAAGA--CGT--AT-GGACCCGT-GCGCT--TAATGGTG--CATT-TCACGCGAGA-AACGGAGGCAG-GTGCCAT--CCGCG-GC----GAAAAGATCATCAGACATGACATAAC-
GAGGAAACCGAT-CCGGAGTGAATACGTC--ACATG-CCAGGGG-GCGGTGTGCAG-TTCTC-GACCG--GAAACGCCCACCC--ACGACGTACCG-CCAAAGCGTCGCCTGG
Score     :  647.5
```

5. test5.seq:

```
Enter the name of the input File: test5.seq

********** Result **********

Sequence 1:  AGGC--CGAAAACGTCGCGAAT-T-GACCCTGGCGACGCCGCCGAACGGGACCTCCGTTAGT-G---T-GGGAGGTCATCAATCTCGTT---CGCTAGCGGCTGACACCAATCACTATAAGTCTGTCATGAC
Sequence 2:  ---CTTC---AA-GTCA--AATATAGATCCTGGC--CGCT-CC--ACGGG--CT---TAAGTCGTTCTCCGAAGGT-A-CGATCTGGTTGGATGCTTCCGTCTAA-AC-AAGAAG-ATAA-----TC--G--
Score     :  92.0
```