项目一：MSA

人工智能　　CS410　　2021年秋季

姓名：李子龙　学号：518070910095　日期：2021 年 10 月 10 日

目录

1 题目

1.1 Topic

Implement three algorithms to solve multiple sequence alignment (MSA) problems.

1.2 Requirements

(1) Implement dynamic programming (DP) algorithm to find the optimal solution.
(2) Implement A-star (A*) algorithm to find the optimal solution.
(3) Implement genetic algorithm to find the optimal/suboptimal solution.

1.3 Rules

表 1: Cost Matrix

	Match α_{pp}	Mismatch α_{pq}	Gap δ
Cost	0	3	2

The table above shows the pairwise cost matrix. For multiple sequence alignment, the cost should be calculated in a cycle pairwise manner. Note that GAP-GAP is a match and should be considered as 0 cost. For every query, find the best alignment(s) in the database with the lowest cost.

2 动态规划算法

在算法与复杂性课程[1]里，已经提到了双序列比对的动态规划算法。

Algorithm 1: 双序列比对动态规划 MSA

Input: $x_1x_2\cdots x_m, y_1y_2\cdots y_n, \alpha, \delta$
Output:

1 **for** $i \leftarrow 0$ *to* m **do** $M[i,0] = i\delta$;
2 **for** $j \leftarrow 0$ *to* n **do** $M[0,j] = j\delta$;
3 **for** $i \leftarrow 1$ *to* m **do**
4 　　**for** $j \leftarrow 1$ *to* n **do**
5 　　　　$M[i,j] = \min(\alpha[x_i,y_j] + M[i-1,j-1], \delta + M[i-1,j], \delta + M[i,j-1])$;
6 **return** $M[m,n]$;

参考文献

[1] XIAOFENG G. Algorithm & complexity class lab 06[EB/OL]. 2021. https://github.com /LogCreative/AlgAndComplexity/blob/master/Lab06/Code-SequenceAlignment.cpp.