

Sequence Alignment 보고서

201504280

컴퓨터공학과 신윤호

1. 구현 내용

1) 초기화

먼저 stringInput이라는 메소드를 정의해서 표준 입력을 통해 두 개의 문자열을 입력받도록 구현했다. 각각의 문자열은 StringA, StringB 변수에 저장된다. 그 후, 각 문자열의 길이를 토대로 OPT테이블을 생성했다.

한편 gap 비용은 1, mismatch 비용은 2로 고정되어 주어졌으므로 이를 각각 변수 상수인 GAP_COST, MISMATCH_COST에 저장했다.

2) runSequenceAlignment

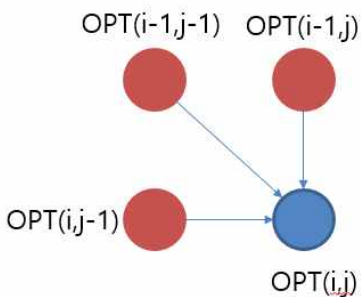
sequence alignment 알고리즘의 본체 함수다. 전반적으로 교재의 점화식을 그대로 코드로 구현했다. 먼저 for문을 이용해 0번째 행의 요소들을 초기화하고, 한번더 for문을 사용해 각 행의 0번째 열을 초기화해 주었다. 이는 교재에 제시된 점화식에서 $j\delta$ (if $i=0$), $i\delta$ (if $j=0$)를 구현했다고 할 수 있다.

다음으로 나오는 for문에서는 3가지 경우중 하나가 선택된다. 첫 번째 경우는 $x_i y_j$ 를 align시키는 경우로써 이 경우에는 $\alpha_{x_i y_j} + OPT(i-1, j-1)$ 이 cost가 된다. $\alpha_{x_i y_j}$ 의 값은 mismatch cost가 2로 고정되어 주어졌으므로 단순히 두 문자가 일치할 경우에는 0이 되고 일치하지 않을 경우에는 MISMATCH_COST를 가지도록 하였다. 두 번째 경우는 x_i 를 gap시키는 경우다. 이 경우에는 $\delta + OPT(i-1, j)$ 를 하도록 구현했다. 마지막 경우는 y_j 를 gap시키는 경우다. 이 경우는 $\delta + OPT(i, j-1)$ 을 하도록 구현했다.

어떤 케이스를 선택할지 여부는 세 경우의 값들을 각각 구해, 그중에서 최소값이 되는 경우를 선택하도록 했다.

3) printTable

MIN COST 및 OPT테이블, 그리고 경로를 출력하는 함수다. MIN COST는 단순히 OPT테이블에서 얻어진 optimal한 값을 출력하도록 했다. OPT테이블 또한 runSequenceAlignment에서 만들어진 OPT테이블을 출력하도록 했다. 경로를 추적하는 부분이 구현하기 힘들었는데, 수업 시간에 sequence alignment 경로를 그래프 최단거리 구하듯 구할 수 있다고 하였던 말씀이 생각나 다음 그림과 같은 상황을 가정해서 구현했다.



$OPT(i, j)$ 는 위 그림과 같이 세 방향 중 한 방향에서 온다. 때문에 진행 방향이 될 수 있는 경로는 왼쪽, 위쪽, 대각선 위쪽 방향이라고 생각했다. 세 경우 중 진행 방향을 선택하는 것은 cost값이므로 각각의 방향으로 가는 cost를 구해서 그중 값이 최소인 것을 선택하도록 했다. 즉 각각의 cost값은 weight로 보고 가중치 값이 최소인 노드를 선택하도록 구현했다. 한편 cost를 구하는 방법은 sequence alignment 알고리즘에서 cost구하는 방법을 그대로 가져다 썼다. 가령 대각선 방향의 cost는 $\alpha_{x_i y_j} + OPT(i-1, j-1)$ 로 구했다.

2. 결과 화면

"C:\Program Files\Java\jdk1.8.0_121\bin\java" ...

String 1 : occurrence

String 2 : occurrence

MINCOST : 3

		o	c	c	u	r	r	e	n	c	e
	0	1	2	3	4	5	6	7	8	9	10
o	1	0	1	2	3	4	5	6	7	8	9
c	2	1	0	1	2	3	4	5	6	7	8
u	3	2	1	2	1	2	3	4	5	6	7
r	4	3	2	3	2	1	2	3	4	5	6
r	5	4	3	4	3	2	1	2	3	4	5
a	6	5	4	5	4	3	2	3	4	5	6
n	7	6	5	6	5	4	3	4	3	4	5
c	8	7	6	5	6	5	4	5	4	3	4
e	9	8	7	6	7	6	5	4	5	4	3

OPT[9][10] -> OPT[8][9] -> OPT[7][8] -> OPT[6][7] -> OPT[6][6] -> OPT[5][6] -> OPT[4][5] -> OPT[3][4] -> OPT[2][3] -> OPT[2][2] -> OPT[1][1] -> OPT[0][0]

Process finished with exit code 0

"C:\Program Files\Java\jdk1.8.0_121\bin\java" ...

String 1 : CTGACCTACCT

String 2 : CCTGACTACAT

MINCOST : 4

		C	C	T	G	A	C	T	A	C	A	T
	0	1	2	3	4	5	6	7	8	9	10	11
C	1	0	1	2	3	4	5	6	7	8	9	10
T	2	1	2	1	2	3	4	5	6	7	8	9
G	3	2	3	2	1	2	3	4	5	6	7	8
A	4	3	4	3	2	1	2	3	4	5	6	7
C	5	4	3	4	3	2	1	2	3	4	5	6
C	6	5	4	5	4	3	2	3	4	3	4	5
T	7	6	5	4	5	4	3	2	3	4	5	4
A	8	7	6	5	6	5	4	3	2	3	4	5
C	9	8	7	6	7	6	5	4	3	2	3	4
C	10	9	8	7	8	7	6	5	4	3	4	5
T	11	10	9	8	9	8	7	6	5	4	5	4

OPT[11][11] -> OPT[10][10] -> OPT[10][9] -> OPT[9][9] -> OPT[8][8] -> OPT[7][7] -> OPT[6][6] -> OPT[5][6] -> OPT[4][5] -> OPT[3][4] -> OPT[2][3] -> OPT[1][2] -> OPT[1][1] -> OPT[0][0]

Process finished with exit code 0