NATIONAL INSTITUTE OF TECHNOLOGY, WARANGAL

Dynamic Band Approach for LCS Calculation of Two Sequences

Asankhaya Sharma and Govindarajan S
Department of Computer Science and Engineering, NIT Warangal

Abstract

The problem of finding longest common subsequence has many applications in bioinformatics. In this paper we present a new method for calculating the longest common subsequence of two strings based on parameterized computation, by dynamically changing the value of the parameter. This algorithm performs far better then the usual dynamic programming based methods and other parameterized approach. Since the algorithm changes the value of the parameter dynamically, theoretically its the only method that can give the optimal result. This approach works by reducing the sets of cases to be considered in the dynamic programming based method, the time complexity achieved is O(cn). There is no other method which gives this fast response time.

Introduction to the problem of finding LCS of two Sequences

The problem of finding the longest common subsequence(LCS) for two sequences has got a wide range of applications. One such application is in sequence comparison(eg. Multiple sequence alignment) in DNA (represented by four-letter alphabets) and protein(represented by twenty-letter alphabets) sequence A string "k" is a subsequence of some string "s" if the deletion of some characters in "s" yields the string "k". Thus the longest common subsequence of two strings is a common subsequence of maximum length. For e.g. if x=" abccbc" and y="acbbc" then the longest common subsequence of the two strings is "acbc" .As we have already mentioned the solution for this problem is critical for its application in bioinformatics.

Dynamic Band Algorithm

The dynamic programming bases approach has a time complexity of O(n^3). While using the parameterized method for LCS calculation we can improve the time bound. This algorithm modifies the parameterized approach by using the value of the band dynamically at runtime. We can choose the value according to the expected string properties to get an even better time bound. The following code illustrates a simple instance when the band is incremented after each iteration.

```
LCS()
{
//x and y are the two strings
// L is the two dimensional array used in the dynamic
// programming method

do{

      for(int i=0;i<x.length()+1;i++)
         L[i][0]=0;
      for(int i=0;i<y.length()+1;i++)
         L[0][i]=0;
      for(int i=1;i<x.length()+1;i++)
      {
         for(int j=1;j<y.length()+1;j++)
          if(i==j || (i>j && i-j<=band) ||(j>i && j-i<=band))
          {
                        if(x.charAt(i-1)==y.charAt(j-1))
                        L[i][j]=L[i-1][j-1]+1;
                        else if(L[i-1][j]>=L[i][j-1]
                        L[i][j]=L[i-1][j];
                        else L[i][j]=L[i][j-1];
           }
           band++;
       }
}while(LCS2(x,y)!=L[x.length()][y.length()]);
}
```

Value of the band	Dynamic Programming	Parameterized Method	Our Algorithm
	Time Needed in Seconds		
100	0.21	0.14	0.13
200	1.10	0.19	0.15
500	6.10	0.23	0.19
1000	11.02	0.28	0.22

Other Approaches

Dynamic programming

```
m = length[X]
n = length[Y]
for i = 0 to m do c[i, 0] = 0
for j = 0 to n do c[0, j] = 0
for i = 1 to m do for j = 1 to n do
  if xi = yj
  then c[i, j] = c[i-1, j-1]+1
  else if c[i-1, j] >= c[i, j-1]
       then c[i, j] = c[i-1, j]
       else c[i, j] = c[i, j-1]
return c and b
```

Parameterized Approach

```
m = length[X]
n = length[Y]
for i = 0 to m do c[i, 0] = 0
for j = 0 to n do c[0, j] = 0
for i = 1 to m do for j =1 to n
  if ((i, j) is within BAND) {
  do if xi = yj
    then c[i, j] = c[i-1, j-1]+1
    else if c[i-1, j] >= c[i, j-1]
    then c[i, j] = c[i-1, j]
    else c[i, j] = c[i, j-1]
}
```

References

-- J. K. Lanctot, M. Li, B. Ma, S. Wang, and L. Zhang, "Distinguishing string selection problems," Infmation and Compuation., vol. 185, No. 1, pp. 41-55, 2003.

-- M. Li, B. Ma, and L.Wang, "On the closest string and substring problems," Jounal of the ACM, vol. 49, pp. 157-171, 2002.

[13] J. W.

-- R. Downey and M. Fellows, Parameterized Complexity, Springer, New York, 1999.

-- Yuan Lin, Jeff Jenness and Xiuzhen Huang, Proceedings of the International Conference of Bioinformatics and Computation 2006.

Contact Information

Asankhaya Sharma
asankhaya@yahoo.com
Govindarajan S
hari_9766@yahoo.co.in