CSE 373: Analysis of Algorithms

Assignment 5 (Dec. 1st 2016)

Due date and time: Dec 14th 12:00p.m. (noon)
Submit in my office B422 (handwritten hardcopy).
You can work in pairs.

1. [10] Implement the dynamic programming algorithm for approximate string matching (in whatever language you wish) and use it to find the best alignment between the following pairs of strings:

a) “watch the movie raising arizona?", “watch da mets raze arizona?"

b) “this is what happens when I type slow", “htishisth whaty havpens when ui type fasht"

c) “leonard skiena", “lynard skynard"h

* Submission Guideline for Problem 1: Hardcopy of your code, the matrixes generated for a) b) and c), and the alignment result for a) b) and c). Also, write down the recurrence function and cost for substitution, deletion, and insertion.

2. *[6]* In the United States, coins are minted with denominations of 1, 5, 10, 25, and 50 cents. Now consider a country whose coins are minted with denominations of $\{d1, \ldots, dk\}$ units. We want to count how many distinct ways $C(n)$ there are to make change of n units. For example, in a country whose denominations are $\{1, 6, 10\}$, $C(5) = 1$, $C(6)$ to $C(9) = 2$, $C(10) = 3$, and $C(12) = 4$.

(a) How many ways are there to make change of 20 units from $\{1, 6, 10\}$?

(b) Give an efficient algorithm to compute $C(n)$, and analyze its complexity.
(Hint: think in terms of computing $C(n, d)$, the number of ways to make change of n units with highest denomination d. Be careful to avoid overcounting.)

3. *[6]* The traditional world chess championship is a match of 24 games. The current champion retains the title in case the match is a tie. Each game ends in a win, loss, or draw (tie) where wins count as 1, losses as 0, and draws as 1/2. The players take turns playing white and black. White has an advantage, because he moves first.
The champion plays white in the first game. He has probabilities ww, wd, and wl of winning, drawing, and losing playing white, and has probabilities bw, bd, and bl of winning, drawing, and losing playing black.

(a) Write a recurrence for the probability that the champion retains the title. Assume that there are g games left to play in the match and that the champion needs to win i games (which may end in a 1/2).

(b) Based on your recurrence, give a dynamic programming to calculate the champion’s probability of retaining the title.

(c) Analyze its running time for an n game match.

4. *[5]* Consider the problem of storing n books on shelves in a library. The order of the books is fixed by the cataloging system and so cannot be rearranged. Therefore, we can speak of a book b_i, where $1 \le i \le n$, that has a thickness t_i and height h_i. The length of each bookshelf at this library is L. Suppose all the books have the same height h (i.e. , $h = h_i = h_j$ for all i, j) and the shelves are all separated by a distance of greater than h, so any book fits on any shelf. The greedy algorithm would fill the first shelf with as many books as we can until we get the smallest i such that b_i does not fit, and then repeat with subsequent shelves. Show that the greedy algorithm always finds the optimal shelf placement, and analyze its time complexity. (Hint: use proof by contradiction)

5. *[7]* This is a generalization of the previous problem. Now consider the case where the height of the books is not constant, but we have the freedom to adjust the height of each shelf to that of the tallest book on the shelf. Thus the cost of a particular layout is the sum of the heights of the largest book on each shelf.

(a) Give an example to show that the greedy algorithm of stuffing each shelf as full as possible does not always give the minimum overall height.

(b) Give an algorithm for this problem, and analyze its time complexity. Hint: use dynamic programming.

6. *[3]* Give the 3-SAT formula that results from applying the reduction of SAT to 3-SAT for the formula:

$$(x \lor y \lor \bar{z} \lor w \lor u \lor \bar{v}) \land (\bar{x} \lor \bar{y} \lor z \lor \bar{w} \lor u \lor v) \land (x \lor \bar{y} \lor \bar{z} \lor w \lor u \lor \bar{v}) \land (x \lor \bar{y})$$

7. *[4]* Draw the graph that results from the reduction of 3-SAT to vertex cover for the expression

$$(x \lor \bar{y} \lor z) \land (\bar{x} \lor y \lor \bar{z}) \land (\bar{x} \lor y \lor z) \land (x \lor \bar{y} \lor \bar{x})$$

8. *[5]* The *baseball card collector problem* is as follows. Given packets *P*1, . . . , *Pm*, each of which contains a subset of this year's baseball cards, is it possible to collect all the year's cards by buying $\leq k$ packets?
For example, if the players are *{Aaron,Mays,Ruth, Skiena}* and the packets are *{{Aaron,Mays}, {Mays,Ruth}, {Skiena}, {Mays, Skiena}},* there does not exist a solution for $k = 2$, but there does for $k = 3$, such as *{Aaron,Mays}, {Mays,Ruth}, {Skiena}*
Prove that the baseball card collector problem is NP-hard using a reduction from vertex cover.

9. *[5]* Show that the following problem is NP-complete:
Problem: Dense subgraph
Input: A graph *G*, and integers *k* and *y*.
Output: Does *G* contain a subgraph with exactly *k* vertices and at least *y* edges?

10. *[6]* An *Eulerian cycle* is a tour that visits every edge in a graph exactly once. An *Eulerian subgraph* is a subset of the edges and vertices of a graph that has an Eulerian cycle. Prove that the problem of finding the number of edges in the largest Eulerian subgraph of a graph is NP-hard. (Hint: the Hamiltonian circuit problem is NP-hard even if each vertex in the graph is incident upon exactly three edges.)

11. *[6]* Prove that the following problem is NP-complete:
Problem: Hitting Set
Input: A collection *C* of subsets of a set *S*, positive integer *k*.
Output: Does *S* contain a subset *S'* such that $|S'| \leq k$ and each subset in *C* contains at least one element from *S'?*