CSE 2320 Notes 6: Greedy Algorithms

(CLRS 16.1-16.3)

6.A. Concepts

Commitments are based on local decisions:

- NO backtracking (will see in stack rat-in-a-maze - Notes 10)
- NO exhaustive search (will observe in dynamic programming - Notes 7)

Approaches:

1. Sort all items, then make decisions on items based on ordering.
2. Items are placed in heap and then processed by loop with delete and priority changes.

MAIN ISSUE: NOT efficiency . . . Quality of Solution instead

Special situations - exact solution (these three path problems are asides for now . . .)

Prim’s Minimum Spanning Tree (Notes 15, min-heap)

\[ n \text{ vertices} - \text{choose} \ n - 1 \text{ edges to give tree with minimum sum of (undirected) edge weights.} \]

Path for each vertex is one that minimizes the maximum weight appearing on the path.

Each vertex is labeled with its predecessor on path back to the source (vertex 0).

Each round augments the tree with the minimum weight edge.

So, vertices are finalized in ascending “min of maxes” order \((0, 3, 6, 7, 2, 1, 5, 4)\).
Dijkstra’s (http://www.cs.utexas.edu/~EWD) Shortest Path (Notes 16, min-heap)

- Choose \( n - 1 \) edges to give tree with a path from source to each vertex that minimizes the sum of (directed) edge weights on the path.

- Each vertex is labeled with its shortest path distance from source and its predecessor.

- Each round augments the tree with the last edge for the shortest (uncommitted) path.

So, vertices are finalized in ascending shortest-path distance order \((0, 3, 7, 2, 1, 5, 4, 6)\).

Maximum Capacity Path for Network Flow (Notes 17, max-heap)

- Choose \( n - 1 \) edges to give tree with a path from source \( (0) \) to each vertex that maximizes the minimum capacity of the (directed) edge weights on the path.

- Each vertex is labeled with its maximum capacity from source and its predecessor.

- Each round augments the tree with the last edge for the maximum capacity (uncommitted) path.

So, vertices are finalized in descending maximum capacity order \((0, 1, 2, 6, 7, 3, 4, 5)\).

More frequently - heuristic (approximation)

6.B. EXAMPLE – activity scheduling (unweighted interval scheduling)

- \( n \) activities

- Start time (activity starts \textit{exactly} at time)

- Finish time (activity finishes \textit{before} this time)

One room
Goal: Maximize number of activities. (Unlike weighted interval scheduling in Notes 7)

Greedy Solution:

1. Sort activities in ascending finish time order.

2. Consider each activity according to sorted order:

   - Include activity in schedule only if it does not overlap with other activities already in schedule

*Optimal or heuristic?*

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>D</th>
<th>G</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>E</td>
<td>F</td>
<td>H</td>
<td>J</td>
</tr>
</tbody>
</table>

Optimality Proof:

1. Suppose there is an alternate (optimal) schedule with a different first activity:

   \( s_? \ldots f_? < \text{rest of schedule} > \)

   But \( s_1 \ldots f_1 \) can replace \( s_? \ldots f_? \), since \( f_1 \leq f_? \)

2. Same argument applies to replacing other activities in the schedule

*Problems that can be solved optimally by a greedy method have a simpler structure than problems requiring dynamic programming.*

6.C. KNAPSACK PROBLEM

Can carry \( k \) pounds (to sell) in your knapsack.

Wish to maximize the amount of revenue.

Greedy approach: Choose according to descending order of $$$/lb.
Fractional (divisible) version:

\$$/lb for each divisible item.

Example:

\[ k = 10 \text{ lbs} \]

- **Perfume:** $500/lb, 1 lb available
- **Chocolate:** $30/lb, 5 lbs available
- **Beans:** $2/lb, 5 lbs available
- **Rice:** $1/lb, 5 lbs available

*Optimal or heuristic?*

0/1 (indivisible) version:

Example:

\[ k = 10 \text{ lbs} \]

- **Bottle of wine:** 5 lbs, $100 ($20/lb)
- **Rare book:** 6 lbs, $102 ($17/lb)
- **Sword:** 4 lbs, $76 ($19/lb)
- **Lobster** 2 lbs, $42 ($21/lb)

Greedy says to choose _______________, but optimal is ________________.
6.D. **HUFFMAN CODES** - elementary data compression for a *static* distribution of symbols in an *alphabet*.

Prefix Code Tree

[Diagram of a HUFFMAN CODE tree showing the nodes and paths for symbols A, B, C, D, E, and F, with binary codes 01000000001111001001, and the expected bits per symbol calculated as Σ Probability•Bits = Σ=2.95 = Expected bits per symbol.]

**Concept**: Letters that appear more often (higher probability) should be assigned shorter codes.

**Evaluating a particular code tree (even if not optimal)**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Probability</th>
<th>Bits</th>
<th>Probability•Bits</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>.2</td>
<td>2</td>
<td>.4</td>
</tr>
<tr>
<td>B</td>
<td>.05</td>
<td>3</td>
<td>.15</td>
</tr>
<tr>
<td>C</td>
<td>.3</td>
<td>4</td>
<td>1.2</td>
</tr>
<tr>
<td>D</td>
<td>.15</td>
<td>4</td>
<td>.6</td>
</tr>
<tr>
<td>E</td>
<td>.1</td>
<td>2</td>
<td>.2</td>
</tr>
<tr>
<td>F</td>
<td>.2</td>
<td>2</td>
<td>.4</td>
</tr>
</tbody>
</table>

Σ=1.0   Σ=2.95 = Expected bits per symbol

**Algorithm**: Build up subtrees by pairing trees with lowest probabilities (use min-heap).

[Diagram of the construction of the HUFFMAN CODE tree by pairing the nodes with the lowest probabilities.]

Very easy to implement tree using table with $2n-1$ entries (http://ranger.uta.edu/~weems/NOTES2320/huffman.c):

<table>
<thead>
<tr>
<th>i</th>
<th>probability</th>
<th>left</th>
<th>right</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>.05</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>.3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>.15</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>.1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>.2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>.15</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>.3</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>.4</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>.6</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>1.0</td>
<td>8</td>
<td>9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Probability</th>
<th>Bits</th>
<th>Probability•Bits</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>.2</td>
<td>2</td>
<td>.4</td>
</tr>
<tr>
<td>B</td>
<td>.05</td>
<td>4</td>
<td>.2</td>
</tr>
<tr>
<td>C</td>
<td>.3</td>
<td>2</td>
<td>.6</td>
</tr>
<tr>
<td>D</td>
<td>.15</td>
<td>3</td>
<td>.45</td>
</tr>
<tr>
<td>E</td>
<td>.1</td>
<td>4</td>
<td>.4</td>
</tr>
<tr>
<td>F</td>
<td>.2</td>
<td>2</td>
<td>.4</td>
</tr>
</tbody>
</table>

\[\sum=1.0\quad\sum=2.45=\text{Expected bits per symbol}\]

Optimality: If two minimum-weight trees are not the ones combined, then the expected bits per symbol will be larger than would be computed by the algorithm.

Time: If there are $n$ symbols, then there are $n-1$ subtree combining steps to perform. Each step calls heapExtractMin twice and minHeapInsert once. $O(n \log n)$ overall.

(Aside, more in Notes 7) - Ordinary Huffman coding is not order preserving. The result of comparing two strings, before and after compression, may be different.

Under what condition will a Huffman code tree be order preserving?