C950 Documentation

The Dynamic Algorithm

The algorithm I used for my project is commonly known as a dynamic approach algorithm. The basic idea for this algorithm is to break down a large task into many smaller tasks that are individually much faster and easier to complete. Effectively, to determine the best list for directions, my algorithm looks for the shortest distance between two of the addresses, and pairs them together. It continues to pair up the shortest distances until it can complete a route through all addresses.

Overview of the Program

Algorithm Pseudocode

```
Input: a list of addresses, along with distance matrices for all other addresses
List of addresses = inList
Distance matrix = dDict
pairWithShortest() function (Finds shortest other address, and pairs as a list)
While len(inList) > 1:
    For each in inList:
        newPair = pairWithShortest(each)
        inList.append(newPair)
        inList.remove(each, newPair[-1])
Return inList[0]
```

Programming Environment

The programming environment used for this project was Visual Studio Code, version 1.65, with a 64-bit python version 3.9.10. I personally prefer this software due to its flexibility and debugging features. It also hosts a large variety of useful plugins for most languages.

This was run on a machine running windows 10.

Complexity of the program

Individual Functions:

- dataReset Due to the function only being used to clear a number of lists, the big O notation is O(n)
- insertPackage Due to the nature of appending to a list, this function has a complexity of O(1)
- updatePackage Searching and updating a singular item in a list has a complexity of O(n)
- returnPackage Searching the list has a complexity of O(n)
- dynamicProgrammingApproach Due to the looping nature of the program, the time complexity of the program is O(n^3), which while better than brute force O(n!), likely has plenty of room for improvement.
- addressIdFind Search algorithm for a list O(n)
- addressFind Search algorithm for a list, but with a different search parameter O(n)
- load Addresses due to multiple nested loops, the big O of this one is O(n^2)
- load_Packages no nested loops, just a single loop for basic searching. Big O: O(n)
- milesToTime basic equation. Big O: O(1)
- timeToMiles basic equation. Big O: O(1)
- distributePackages Due to a loop within a loop nature of this one, it has a complexity of O(n^2)
- findShortestTruck A simple search algorithm O(n)
- givePackageToTrucks a single loop, thus simply O(n)
- deliverPackagesOnRoute By far the least efficient function. 3 loops nested in a loop.
 O(n^4)
- Main Technically has a big O of O(n) due to a loop, but as this is the GUI, it is unlikely to actually take anything like that amount of time.

The whole program:

The entire program has 3 main segments:

- 1. A search function for individual packages. This function, as stated before, has a big O average of O(n), as it is simply searching down a list and returning what it finds.
- 2. A return all packages function to display all packages for the day. This function simply prints out the list. As such, it has a big O of O(1)
- 3. An update time function. This updates the entire program, and delivers packages based on the entered time. Due to the deliverPackagesOnRoute function, the big O of this functionality is O(n^4)

Scalability

This program scales reasonably well, though the delivery function could likely be made faster to better increase speed. That said, individual package searches would scale very well. Similarly,

the algorithm used to find the fastest path to each destination is significantly faster than brute force and would scale rather well.

Maintainability

The program has plenty of documentation and commentary in the code itself, to allow for easy navigation and access to the various functions. Similarly, due to the modular nature of the code, debugging problems and upgrading in the future would be very simple.

The self adjusting data structure: Pros and Cons

Pros:

- The hash table used is easily searched, and is readily readable without the need for further machine interpretation.
- Due to the format of the hash table, it would be relatively easy to send reports or similar from data based in it
- There is a secondary data structure to store package information tied to address information. This is used to ensure that no data is lost during processing, and to help optimize the complexity of the program itself

Cons:

- Due to there being some redundant data structures, the space required to store such information is doubled.
- If something is incorrect with one of the data tables, while it is relatively simple to spot that an error has occurred, determining which set to be correct can be made difficult.

Original Program

Attached along with document

The Data Structure

There are four primary data structures used in this program.

- 1. Package Class List
 - a. Used for initial formatting of the package information. Also used to connect packages directly to their address information and the distance matrix
- 2. Address Class List
 - a. Connected to each package is an address class object. This helps organize and optimize the distance matrix used for computation of the algorithm
- 3. Van Class List
 - a. This is used to hold indexes of various packages, along with a total distance driven.
- 4. Package Hash List
 - a. This is used to format the data from the Package class list into a readable and easily searchable format.

Screenshots

Packages at 9:

Packages at 10:

Packages at 1Pm:



Mileage of the trucks after all packages delivered:

Justification of the Core Algorithm

The algorithm I used is a Dynamic algorithm. The primary strengths of this algorithm are its complexity, which is to say that it is relatively simple to understand, and its efficiency, or how close it can get to the best possible solution.

The algorithm easily delivers all packages, with restrictions, before 140 miles is reached.

Two other algorithms I looked at for possible use are the Nearest Neighbor algorithm, which simply looks for the next nearest point until all the points are connected, and the brute force approach.

The nearest neighbor algorithm, while incredibly simple and fast, would likely not have been efficient enough to satisfy all restrictions. As such, I decided against its implementation.

Alternatively, an even simpler algorithm would have been to use brute force to check all possible combinations. While this would have been the best way to find the absolute most efficient route, I'm not positive my computer would have been able to complete the necessary computations before a few months had passed. As such, I decided against using it.

What I would do differently

The main thing I would do differently, is to start smaller. I would bother less with all possible restrictions, and just figure the project out, one step at a time, slowly building on the previous structures with a modular approach.

Justification of Data structures

Verification of requirements

The time needed to complete the lookup function is O(n), due to it being a simple list lookup. As such, time needed to find the appropriate information scales linearly with the total packages in the system. As such, this scales rather well as a data structure.

Due to the redundant nature of the data structure, for every 1 package, 2 packages worth of information is stored. This is not the most efficient use of space, but I believe it to be a worthy trade for the benefits to computation time this gives.

Due to each truck only stores the index of lists, and not lists themselves, the addition of more trucks or addresses would only affect time required for lookup linearly.

Other Data structures that could have been used

A case could be made for Tuples, as they are more space efficient than lists, and due to the nature of my program, tuples could have been used to safely store the information, so long as all package information was formatted and inserted at runtime.

Alternatively, a simply typed array could have been used, which would have provided the most efficient use of space. The tradeoff would be that all information would need to be the same type of data, such as Strings or similar. This would have made the handling of that data quite a bit more tricky on the computational side of things.