# ECS759P Artificial Intelligence: Coursework 1

## Agenda-based Search

### *Implementation details*

The station_dict was converted into a networkx object to be consumed by the BFS and DFS implementations. The code from the lab sessions were modified to fit the London tube map use case.

**State representation:**

BFS and DFS were less demanding on the details for each node, i.e., node_label and parent_node reference would suffice.

Best first search and Depth first search: `[{'label':initial, 'parent':None}]`

In case of Uniform cost search and Uniform cost search extended (including the tube line changes), both demanded more finer details for nodes.

Uniform cost search: `(0, [initial])`

Here, a tuple would hold the cost (average time taken) form the initial node to the current_node and list which would hold the path traversed.

Uniform cost search extended: `[{'label':initial, 'parent':None, 'cost':0, 'line':None}]`

The node/state representation for the Uniform cost search extended, node_label, reference to parent, cost (average time taken) to the current_node from the starting station and finally, to track the tube line switched which will be used to calculate the additional time (2 mins/tube line switched) lost due to change in tube lines throughout the journey.

Best First Search: `(heuristic(start, goal), start, [path])`

Best First search like Uniform cost search extended had an admissible heuristic_cost form a given node to goal node, current_node and a list of stations from starting node to a given node that would represent the path taken.

### *Comparison of BFS, DFS, UCS (refer to table in Appendix section)*

- BFS is known to give the shortest path without additional support from path cost.
- DFS will explore all the nodes along depth due to this we tend to see large node explorations (especially in the inverse explorations) and oftentimes DFS may fail to produce an output due to “infinite” depth explorations.
- UCS is an extension of BFS wherein the path cost is considered for node explorations, i.e., the least cost path is picked in the short term (can result in a high path cost in long term).
- UCS extended is considering the time taken to switch tube lines mid-journey (no pun intended) along with the path cost.
- The node explorations by the above algorithms can loop back to nodes that were already explored, this can be avoided by storing the list of explored nodes and “re-route” to a different node which is not on the list.

### *Conclusion*

BFS can consistently finds the shortest path from the starting station to destination and with least number of node exploration in relative to the other approaches.

***Extending the cost function to incorporate the tube line changes.***

- All the paths see a slight increase in the average time taken to reach the destination due to the added tube line change cost bar Baker Street – Wembley Park as all the stations connecting on this route are on the Metropolitan tube line.
- Breadth first search and Depth first search traversal is unaffected by the tube line changes due its minimal state representation that excludes the path cost and thus the cost incurred on changing the tube line.

# Heuristic search

The heuristic being used can depicted as a piecewise function that follows -

$$H(\text{current\_node}, \text{goal\_node}) = \begin{cases} 0 & \text{current\_node} = \text{goal\_node} \\ 1 & \text{current\_zone} = \text{goal\_zone} \\ |\text{goal\_zone} - \text{current\_zone}| & \text{otherwise} \end{cases}$$

- A zone-based heuristic has been used in Best First Search.
- This heuristic is admissible as absolute difference of the zones numbers is calculated which will not resul in overestimation of the actual cost of the time required to travel form one time zone to the other.
- Less number of nodes will be explored by adding heuristics which is evident from the number of node explorations of Uniform cost search and to that of Best First search (refer to the table in the appendix section). As the heuristic considered above estimates a "metric" (can be time, distance etc.) from the goal node to a given node, cutting down unnecessary node explorations.
- Although average time taken to reach the goal destination may vary from that of Uniform cost search due to alternate path explorations which is restricted only by heuristic (only an estimate) which may not provide all the necessary information (missing path costs) required for an optimal traversal.

***Zone-based heuristic which assigns a constant time of 10 minutes for travel within one zone and a constant time of 20 minutes for travel across zones.***

Advantage –
Since this is a Zone-based heuristic rather than a station-based heuristic, the path to the target may be shorter than with a station-based heuristic.

Disadvantage -
Given that we are only considering the zone-based heuristic, which assigns a constant time of 10 minutes for travel within one zone, there is a good chance that the algorithm will explore nodes within the same zone before reaching the goal, even if the travel time and cost are greater than if it were to follow the actual shortest path.

# Genetic Algorithms

***Implementation details***
A YouTube playlist on genetic algorithms was referred to understand the concepts and was used as a starting ground for the password matching.

***State representation:***
Password consists of random arrangements of uppercase letters, digits and underscore character and is of a fixed length.

***Algorithm components and logic flow:***
- Random passwords were generated for population initialization.
- Fitness scores were calculated for candidates in population and the half of the population was eliminated keeping the candidates with higher fitness scores.
- Two parents are picked out in random form the above list for crossover.
- Multipoint crossover was opted for early convergence directed by a high crossover rate (0.9) producing two offspring.
- The two offspring from the previous step is considered for mutation decided by a low mutation rate (0.1), chosen for early convergence.
- Mutated offspring will now be part of new population for which the above steps are repeated until convergence.

***Experiments by changing Hyperparameters:***
Number of reproductions to converge to the true password taking 10 samples were [28, 25, 24, 28, 27, 24, 27, 22, 24, 27], Mean convergence: 25.6, std of convergence: 1.95 for POPULATION_SIZE = 500, CROSSOVER_RATE = 0.9 and MUTATION_RATE = 0.1.

Slightest change in mutation rate can impact convergence. This was tested by keeping POPULATION_SIZE = 500 and CROSSOVER_RATE = 0.9

| Mutation rate | Mean of Convergence | STD of convergence |
|---|---|---|
| 0.1 | 25.9 | 2.44 |
| 0.2 | 36.0 | 4.35 |
| 0.3 | 73.50 | 15.28 |
| 0.4 | 175.30 | 44.12 |
| 0.5 | 617.70 | 204.59 |

Varying the population size and parametrizing MUTATION_RATE = 0.1 CROSSOVER_RATE = 0.9 results were obtained as below.

| Population size | Mean of Convergence | STD of convergence |
|---|---|---|
| 100 | 48.10 | 16.91 |
| 200 | 36.68 | 6.80 |
| 300 | 29.40 | 2.89 |
| 400 | 25.40 | 4.02 |
| 500 | 24.70 | 3.10 |

# Appendix

*Tabulation of results for all the search algorithms used to search through London tube.*

| Paths | Metrics | DFS | BFS | UCS | UCS Extended | Best FS |
|---|---|---|---|---|---|---|
| **Euston - Victoria** | *Node Explorations* | 25 | 34 | 35 | 30 | 25 |
| | *Avg time taken* | 13 | 7 | 7 | 9 | 15 |
| **Canada Water – Stratford** | *Node Explorations* | 6 | 25 | 59 | 52 | 22 |
| | *Avg time taken* | 15 | 15 | 14 | 18 | 18 |
| **New Cross Gate– Stephney Green** | *Node Explorations* | 32 | 40 | 20 | 18 | 24 |
| | *Avg time taken* | 27 | 14 | 14 | 16 | 22 |
| **Ealing Broadway– South Kensington** | *Node Explorations* | 179 | 47 | 47 | 53 | 11 |
| | *Avg time taken* | 57 | 20 | 20 | 24 | 20 |
| **Baker Street – Wembley Park** | *Node Explorations* | 3 | 9 | 84 | 72 | 6 |
| | *Avg time taken* | 13 | 13 | 13 | 13 | 13 |