Optimization:

The map information from the map is represented in terms the junction links and distance between the junctions. Distribution points and delivery points could be considered as junctions connecting 2 other junctions.

The entire optimization algorithm is broken down into the following steps:

sales_vehicle_alg.png

1. Distribution and delivery point mapping:

The overall objective is to deliver goods from one of the distribution points to the sale points. Therefore for this each need delivery point needs to be assigned with optimal distribution point. Finding the optimal distribution is the objective of this stage.

1. Vehicle and delivery point mapping:

Once the delivery points have been assigned to the distribution points the next task is to assign which vehicle in the distribution centre should deliver to which delivery point. Optimizing this is the objective of this stage.

1. Vehicle delivery sequence:

After assigning some number of delivery points to each vehicle, the task is to identify the optimal sequence to delivery points to cover.

1. Vehicle path planning:

In this stage the path from the source to a destination is found.

Reinforcement learning:

For task 4 an environment is build to interact with an RL agent. The environment is contains the information of the map. The state space is 3Xnumber\_of\_junctions(nj). In this state space 1st nj elements represent the one hot encoded form of the current junction. The next nj elements represents the one hot encoded form of target junction. The last nj elements represent memory that which is 1 for the position of the junction that the vehicle has traversed.

For example ,

let,

nj=5

current position/junction= junction 3

target junction = junction 5

the junctions traversed are 1, 4,3(current position also included)

the current position matrix is:

[0 0 1 0 0]

The target matrix is:

[0 0 0 0 1]

The memory matrix is:

[1 0 1 1 0]

The overall state is:

[0 0 1 0 0 0 0 0 0 1 1 0 1 1 0]

The action space is 3. This represent the 3 other junction that current junction is considered. If the junction is connected to only 2 other junction then it is a not a true junction (pseudo junction) but still works. The 3rd action would be invalid. If it is connected to only one other junction then the other 2 actions are invalid. These 2 and 1 junctions happen at terminal points of the map area or when delivery or distribution points are treated like a junction.

It is important to note that the map parameters are given as inputs to the RL agent. Thus the policy that the agent learns is only suitable to the given particular map.

Reward system:

If the an invalid move is made, that is action where there is no junction, then reward of such action is -10 and the position of the vehicle is not altered.

If the move leads to the target position then the reward is 10 and the episode ends.

If the move leads to an already traversed position then the reward is -10. Since coming back to the same junction would not make sense. There would have been unwanted junctions explored. The episode ends with this action.

If the move leads to a position that is not the target and hasn’t been explored in this episode then the reward is -1(should distance). The negative reward is to penalize long routes.