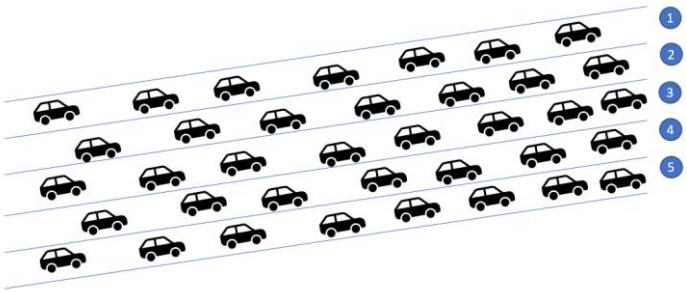
Project

Imagine you are Ah Tan, driving on a long expressway with 5 lanes, and encountered a massive jam. You have to decide whether to keep in your current lane, or to switch to a neighboring lane.

The natural variation in clearance rates (consider this as the number of meters travelled per unit time) dififers across different lanes due to a combination of factors, including:

1. Each vehicle may have different stopping spaces away from the front vehicle. In addition, there can be large heavy vechicles, which takes up more space and at the same time tend to have a lower acceleration.
2. There are other drivers changing lanes. Each driver that switches a lane would lead to one less vehicle for all others below, while at the same time slowly down the lane which he/she switched to.



Vehicles tend to move out from a slower lane and into a faster lane, such that in the next step, the slower lane will speed up slightly while the faster lane will slow down due to more cars in it. We attempt to model this aspect based on the *relative* clearance rate of adjacent lanes.

- For lanes 1 or 5 (side lane), there is only one adjacent lane to compare against. If the side lane has a lower clearance rate than its neighbours (lanes 2 or 4, respectively), 0.2 will be added to the clearance rate of the former and subtracted from that of the latter.
- The converse is true. If the side lane is clearing faster than its neighbor, 0.2 will be subtracted from its clearance rate, and 0.2 will be added to that of its neighbor.
- The effect is cumulative in the middle lanes (2, 3, or 4). If both lanes beside it have a higher clearance rate, 0.4 will be added to its clearance rate. If its clearance rate is higher than only one adjacent lane (ie. faster than one but slower than the other), there is no net change.
- This change is computed concurrently across the lanes.

It is not practical to list everything. For simplicity, we use an uncertainty term N(0,0.1) as well as *random_event* to depict everything outside our consideration. Hence, the environment is governed by the following model, where *v* is initialized independently with uniform probability in [15, 20]:

$$v_{t,j} = v_{t-1,j} + \mathrm{N}(0,0.1) + random\_event + 0.2 \times \mathrm{sgn}(v_{t-1,j-1} - v_{t-1,j}) + 0.2 \times \mathrm{sgn}(v_{t-1,j+1} - v_{t-1,j})$$

If no adjacent lane exists, the sgn() function can be ignored. Note that this transition is part of the environment, and the agent should only be able to observe the events but not be aware of the underlying equation.

Under the *random_event*, there is a 5% chance that the clearance rate is slowed down *by* 20%~50% (uniform probability distribution) of $v_{t-1,j}$, perhaps due to a collision or some obstruction in that lane. There is another 5% chance that the clearance rate is increased *by* 20~40% (also uniform probability distribution) of $v_{t-1,j}$, perhaps due to an ambulance switching on its emergency siren such that vehicles in front make way.

Ah Tan commonly keeps a lookout for 'reference' vehicles (those which are easy to distinguish) in each lane nearby. Every 10 seconds, he takes note of their position, and infers the 'clearance rate' for each of the lanes. This refers to the distance that can be covered within the past 10-second period.

Ah Tan's eyes are very sharp, and he is able to make a judgment with great accuracy. In addition, Ah Tan has psychic powers, and all the vehicles he chose as reference are guaranteed to keep to their lanes. Ah Tan also clearly remembers what happened in the past 30 seconds. Therefore, his decision is based on a list of *three* tuples, where each has *seven* values corresponding to his location d_t away from the destination, his lane l_t, and the clearance rate $v_{t,j}$ for each of the 5 lanes ($j \in \{1,2,3,4,5\}$).
$[(d_{t-2}, l_{t-2}, v_{t-2,1}, v_{t-2,2}, v_{t-2,3}, v_{t-2,4}, v_{t-2,5}),$
$(d_{t-1}, l_{t-1}, v_{t-1,1}, v_{t-1,2}, v_{t-1,3}, v_{t-1,4}, v_{t-1,5}),$
$(d_t, l_t, v_{t,1}, v_{t,2}, v_{t,3}, v_{t,4}, v_{t,5})]$

For example, suppose that at time t, Ah Tan is 4000 meters away from his destination. The state S_t can be something like the following. Notice that the distance covered depends on the clearance rate of whichever lane he is in.
[(4034.7, 3, 18.3, 16.7, 17.9, 17.3, 16.8),
(4017.3, 3, 18.1, 17.1, 17.4, 17.6, 16.9),
(4000.0, 4, 18.2, 17.3, 17.2, 17.3, 17.1)]

The next state, S_{t+1}, is the observation in the next step (ie. 10 seconds later). It could be like:
[(4017.3, 3, 18.1, 17.1, 17.4, 17.6, 16.9),
(4000.0, 4, 18.2, 17.3, 17.2, 17.3, 17.1)
(3982.5, 4, 17.8, 17.5, 17.6, 17.5, 17.2)]

The current distance at S_{t+1} (3982.5, represented in red) is obtained by subtracting the clearance rate (17.5) of lane 4 from the previous distance (4000.0), represented in blue.

Task 1 (Total 25 marks)

Task 1a (7 marks)

Code out the said environment, which is approximated to be in discrete time intervals of 10 seconds apart. Therefore, 10 seconds elapse between each successive action.

The reward can be taken as the distance covered per 10-seconds-period, minus 10 for each time step, in addition to a 'stress' component corresponding to a reward of −5 each time Ah Tan attempts to change lane. Note that there is only a 50% success rate that Ah Tan will be able to change lanes successfully, if he wants to do so. Regardless of success or failure, the reward of −5 is added in that step. If he tried to change lanes but failed, he will remain in that current lane for the 10 second period.

This agent should take in the state as defined in question 1, and return an integer (either -1, 0, or 1) representing to change to the left lane beside it, remain in the current lane, or change to the right lane beside it. If the agent is already at the leftmost lane, and tries to move left (or at the rightmost lane and tries to move right), it will simply stay in its current lane, but the reward of -5 is still added.

Task 1b (3 marks)

Create a rule-based agent to act as a benchmark. Justify, qualitatively and quantitatively, why your group deems this particular agent to be good enough as a benchmark.

Task 1c (10 marks)

Train two distinct RL agents. One of them should involve a policy-based model exclusively. For the other model, you have the choice of using value-based or actor critic.

Submit it as a jupyter notebook showing the training process. Model weights, if relevant, should be submitted as well. Marks for this component will be awarded on a competitive basis, based on the performance relative to that of other groups.

Task 1d (5 marks)

Visualize your rule-based agent and the other two RL agents.

Using matplotlib is sufficient, though you may also make the visualization more fanciful if you wish to. You may use a gif with multiple plotted images, or simply submit a folder of image files labelled according to the time steps.

Note that there is no single 'correct' answer or holy grail for this. Your team may exercise discretion in how you want to answer this question.

Task 2 (Total 15 marks)

Be in the shoes of a teacher now. Create an RL environment which would be structured such that the ideal actions are not immediately obvious and that the effect of actions taken are realized over time.

In your scenario, there would be your car (trained agent) along with 6 other cars. One car would always switch to the fastest lane, while the remaining 5 cars would keep to their respective lanes. Formulate the transition dynamics (which should be stochastic) and train an agent such that your agent has the highest average performance over 500 runs. Note that merely modifying the coefficients of the equation governing $v_{t,j}$ would not get you good marks.