The fluctuating period after the initial learning stage will be referred to ‘confusion stage’ in the following discussion. By observing the Q-table generated and the simulation produced, we can see that there are two main causes for the confusion stage: First, there are some areas in the graph that are harder to traverse than other, such as the corners or the arena, and thus the number of collisions will increase. To increase the agility and allow the vehicle to make more appropriate choices, we need to increase the action space. But 10 more actions means 1000 more entries in the Q-table, and at the beginning of the project we have a state space of 16, we trained our model for 14 hours and the performance still didn’t improve.

However, the second cause has a much more important effect. After training for 1000 to 2000 seconds, the vehicle could have a clear idea about what to do in most of the scenarios, in another word, when to turn right and when to turn left, also how sharp a turn should it make. Then the limiting sensory power started to cause problems. Some scenarios that requires different actions are reflected as the same state, for example heading headfirst towards a wall with a corner nearby requires different actions when the corner is to the left of vehicle than when wall is to the right, but the state will be [0.2 0.2] for both scenarios. This can cause confusions in the Q-table and overwrites some previous choices. But more sensors requires more state. One more sensor will bring 10 times more state, while decreasing the number of states of each sensor can cause the vehicle to loss some of its ability to reason under different scenarios, still causing damage to the obstacle avoidance ability.

The fact that possible increase in vehicle’s performance may bring exponential growth in state and action space is abysmal, and the prospect of wider application is dim.

However, there was a very similar project to ours that manages to achieve some level of success.[[1]](#endnote-1) It shows that this approach can be useful after some more advanced design of the reward signal and it can work with a even simpler state space – their two sensors can only detect whether or not there is an object in range, without actual detailed distance, so two sensors give a state space of 4, which still gives a pretty solid performance in a relatively complex environment. Agility and sensory power can be a problem though if the environment does get more complex, so some more improvements are needed.

Thus, this idea is not entirely unrealistic, but it does requires a lot more care in design than the simplistic design we are working on here.

1. @INPROCEEDINGS{8733621,

   author={T. {Ribeiro} and F. {Gonçalves} and I. {Garcia} and G. {Lopes} and A. F. {Ribeiro}},

   booktitle={2019 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)},

   title={Q-Learning for Autonomous Mobile Robot Obstacle Avoidance},

   year={2019},

   volume={},

   number={},

   pages={1-7},

   doi={10.1109/ICARSC.2019.8733621}} [↑](#endnote-ref-1)