Chart, line chart

Description automatically generatedUsing different values of degree of exploration ‘c’, we can see from figure 1 that as the time step increases, the percentage of correct pulls will follow a logarithmic curve. This is because the rate of increase of upper bound confidence value representing the amount of exploration the algorithm does, diminishes over time since the number of times the charging station has been tried, increases over time. In addition, it can be seen that the rate of convergence for low values of c (less than 1) is faster than high values of c (greater than 1).

Figure 1: graph showing the percentage of correct pulls for different confidence values

The confidence value of less than 1 will give a best performance only if the first charging station chosen was the optimum station because once a charging station is chosen, it will tend to exploit it most however if an incorrect station is first selected, then the performance will be extremely poor. Therefore on average a confidence value of 1 will give the best performance as its balance between exploration and exploitation in most cases will result in a fast conversion on the optimal charging station (figure 1).

Chart, line chart

Description automatically generatedDifferent degrees of exploration rates gives a trade-off between exploration and exploitation. A plot of the variation of the mean total reward vs charging rates (figure 2) shows that the mean total reward fluctuates between high total mean rewards with a peak at around 4.21 and then begins to decrease dramatically when c is greater than 1. This is because the is a representation of the variance so with a c of less than 1, we decrease the variance of charging rates for each charging station by a factor of c and for c greater than 1, the variance of charging rates is increased by a factor of c. This can be seen in figure 1 where higher degrees of exploration takes a longer time to converge on the optimum charging station as the reward at each step has greater variance.

Figure 2: graph showing total mean reward for 200,000 iterations for various confidence levels

Chart, line chart

Description automatically generatedHowever when running the UCB algorithm, we noticed that on many occasion, a value of c less than 0.5 often gave 0% correct action for a lot of steps (figure 3). This is because of the small variance determined by c so if an incorrect charging station was selected first, it would tend to stay with it as the exploitation of the algorithm is high whereas if a correct place was chosen then the result would converge very quickly on the optimum charging station, this is further highlighted in figure 4 which shows what happens to the cumulative regret when selecting an incorrect charging station with low confidence level. When c = 0.1, the regret increases much quicker than when c = 10 because the system was determined to stay with one charging system.

Figure 3: Error case when a confidence level < 1 picks an incorrect action for multiple steps

Chart, line chart

Description automatically generated

Figure 4: cumulative regret for various confidence levels and shows the issue when a low confidence value picks and incorrect arm

This can also work beneficially however if the initial action chosen is optimal as can be seen in figure 5, which shows the arm pulled history over time for confidence levels of 1 and 0.1. With a confidence level of 1, the system openly explores the different charging stations before eventually converging on the optimal solution, whereas confidence level of 0.1, the system exploits one charging station throughout.

Chart

Description automatically generated

Chart, histogram

Description automatically generated

Figure 4: History of used charging stations for confidence levels of 1 and 0.1