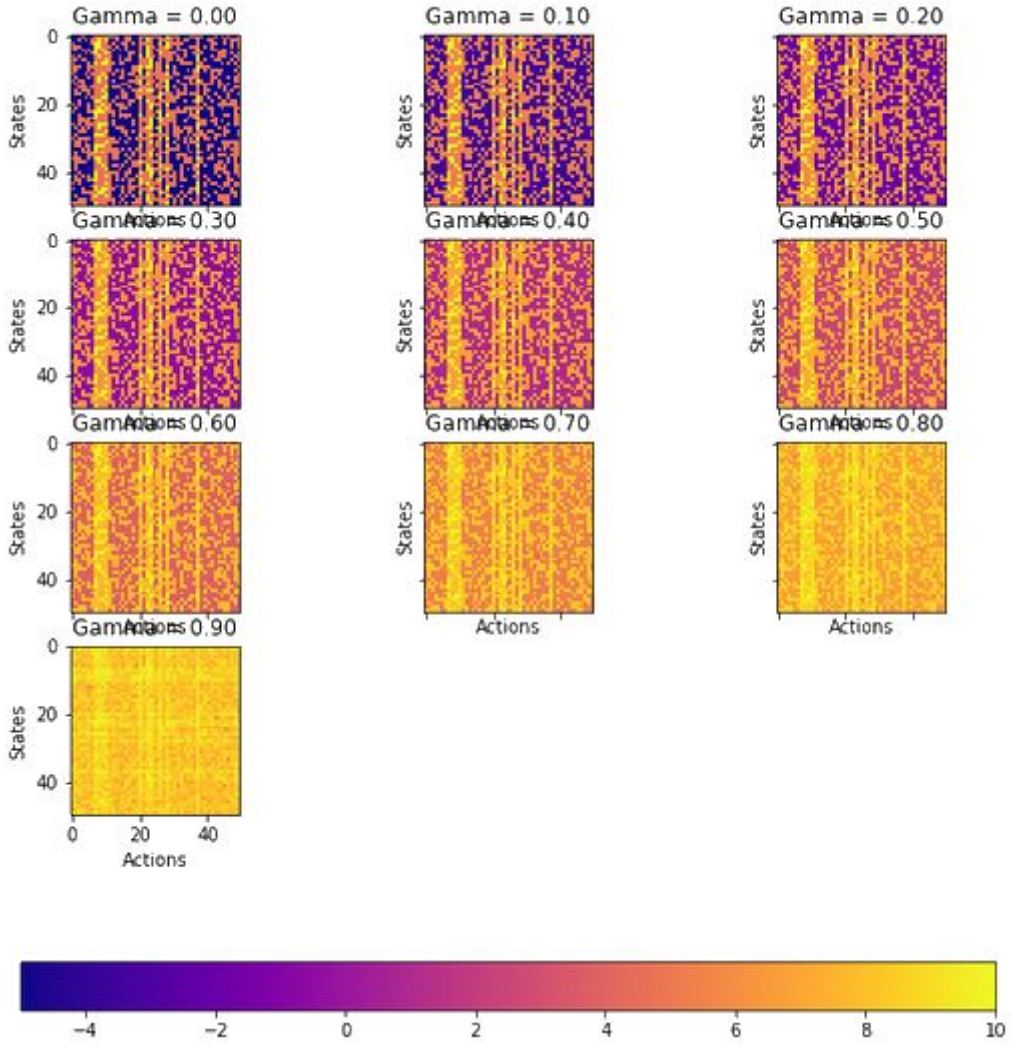
Report 12/05 : Reinforcement learning for Cache-Friendly Recommendations

Jade Bonnet and Clément Bernard

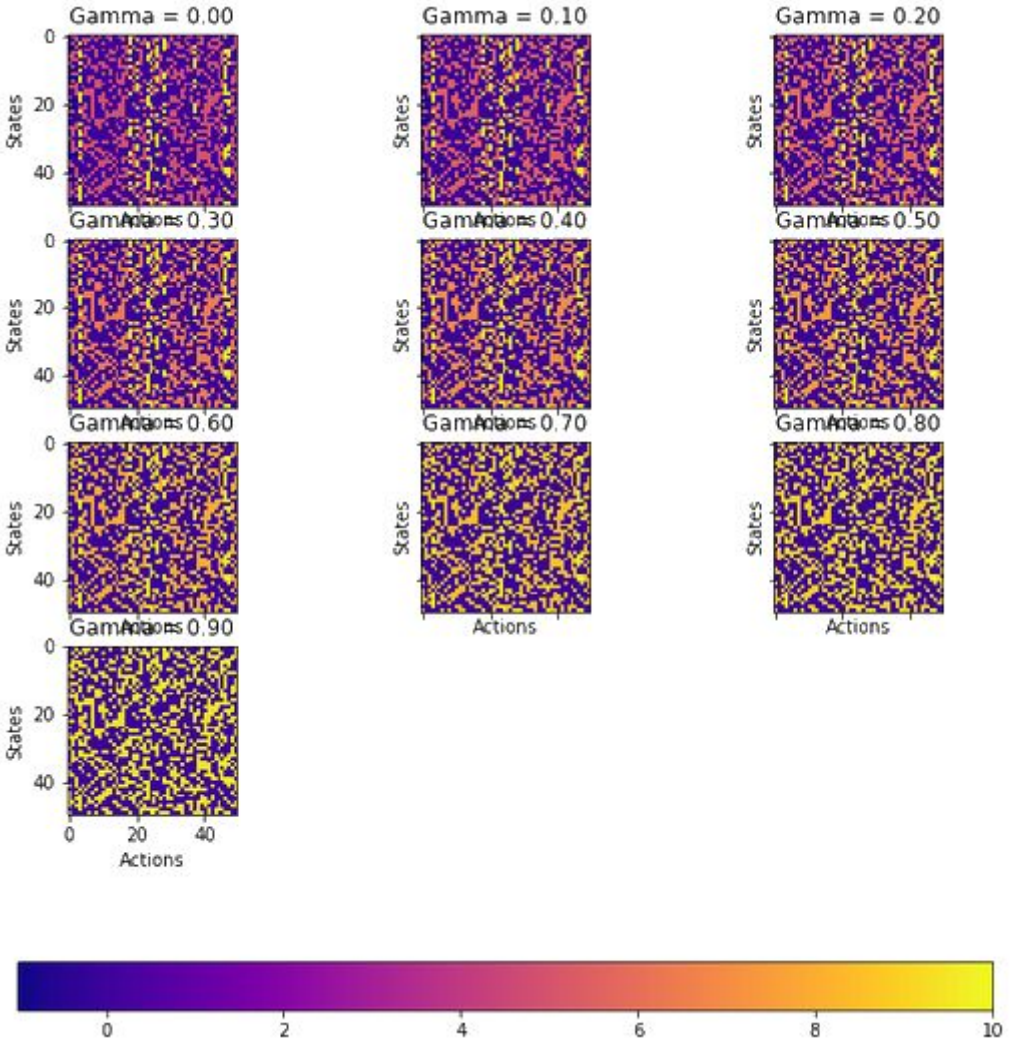
Reminder : we'll say that the algorithm v1 corresponds to the algorithm which explore over the whole catalogue, whereas the algorithm v2 corresponds to the one which explores to the related contents.

1) Normalize the q_table for different values of gamma

We normalized the q_tables by dividing the table by $\frac{1}{1-\gamma}$.
The aim is to see the value per step.



Normalized q_tables for different values of gamma (algorithm v1)

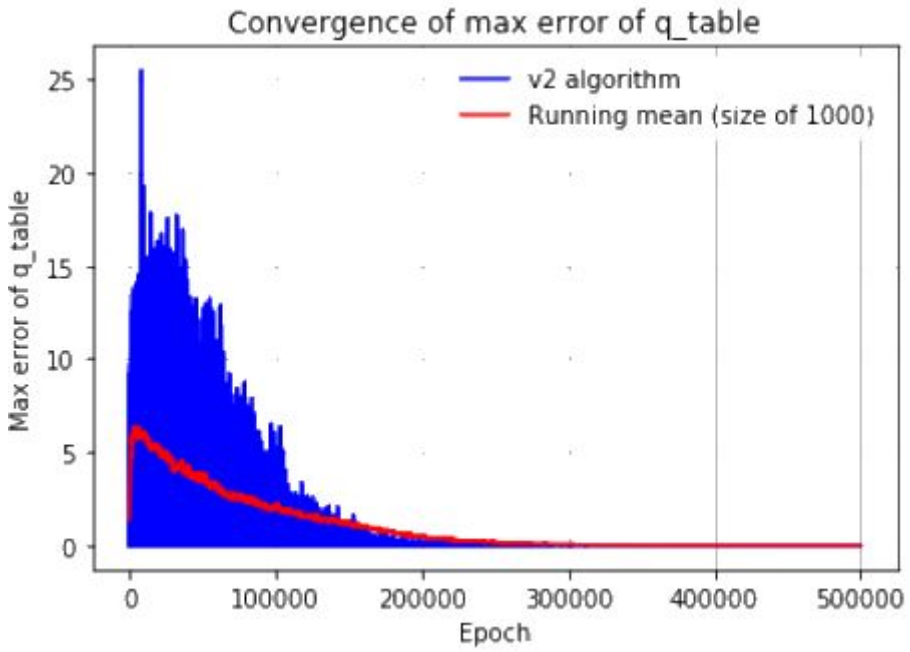


Normalized q_tables for different values of gamma (algorithm v2)

The second plot shows that, for low value of gamma, the agent tends to recommend content with high current rewards whereas for gamma = 0.9, it can recommend contents which don't always lead to high current reward.

2) Find convergence criteria

As suggested, we used the maximum difference between the q_table as a criteria of convergence. Here is the plot of this error depending on the epochs :

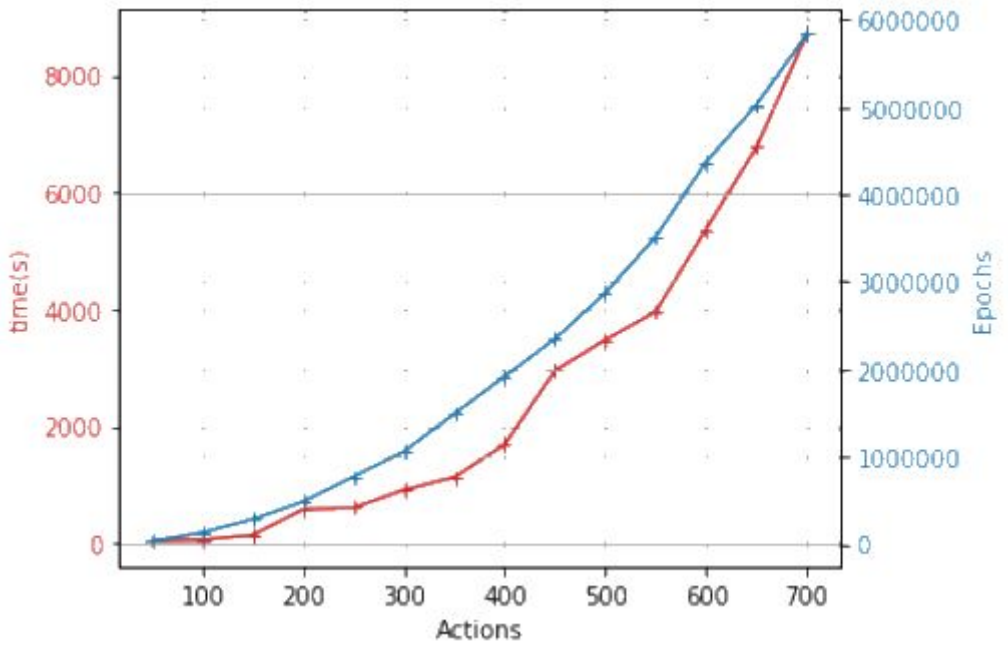


Evolution of the max error of q_table over the epochs

We decided to take as a threshold 0.1.

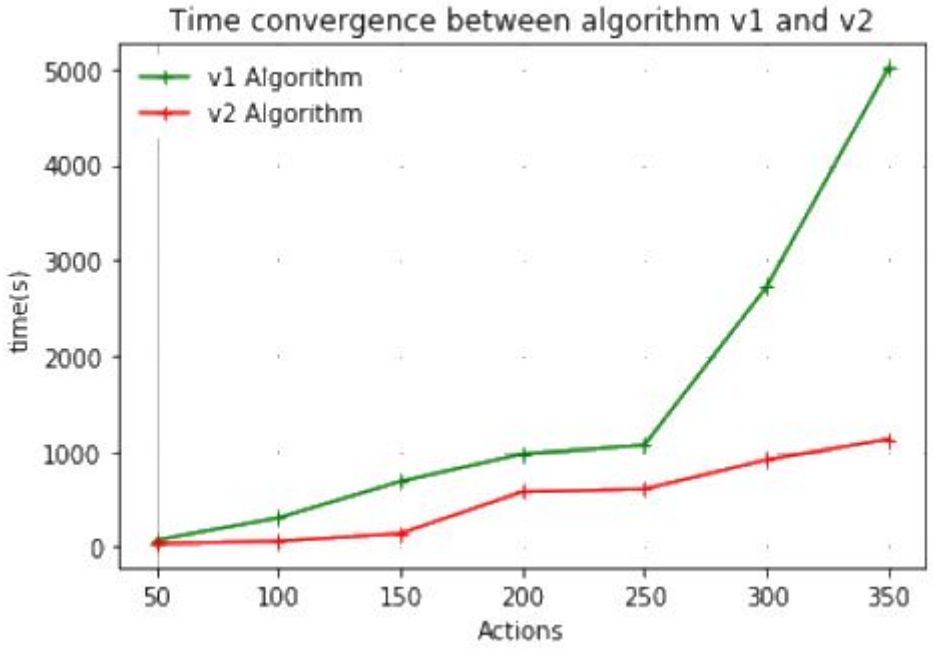
3) Time of convergence

We computed the time to converge and the number of epochs required for each value of gamma. Here is the result :



Time and epoch to converge for different values of gamma (v2 algorithm)

We used the v2 algorithm because it is faster that the first one, as it only explores over the related contents. Here is the comparison between the v1 and v2 algorithm in term of time of convergence :



Time of convergence for algorithm v1 and v2

4) Deep q learning

As Jade didn’t have knowledge on Deep Learning, we decided to split the work by 2.
Jade worked on some tutorials on Deep Learning and Tensor Flow whereas Clement started the tutorials on Deep Q Learning.