CS747 Assignment-3
Windy Gridworld

Rohan Gupta, Roll Number: 180010048

November 2020

Contents

1 Introduction

For all the graphs presented below(except for Sections 3.3 and 4), the values of various parameters are shown in Table 1. In addition, each of the plot and data presented are a result of runs averaged over 50 random seeds.

Parameter	Value
Episodes	200
Learning Rate	0.7
ϵ	0.05

Table 1: Values of parameters

2 Sarsa(0) plots

Figure 1, 2, 3 show the performance of sarsa(0) update on the baseline model, windy gridworld with King's moves, and stochastic wind with King's moves respectively. Stochastic wind is added in the third model, only when there is a non-zero wind present in the cell originally.

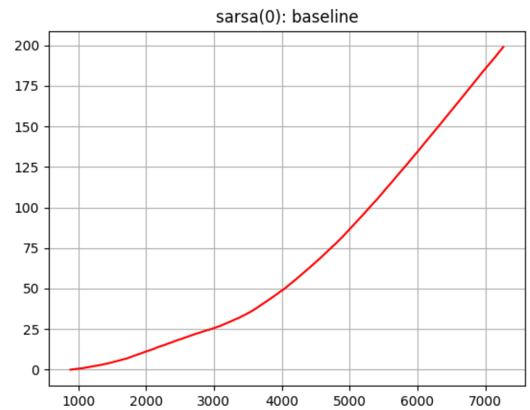


Figure 1: sarsa(0) on baseline model

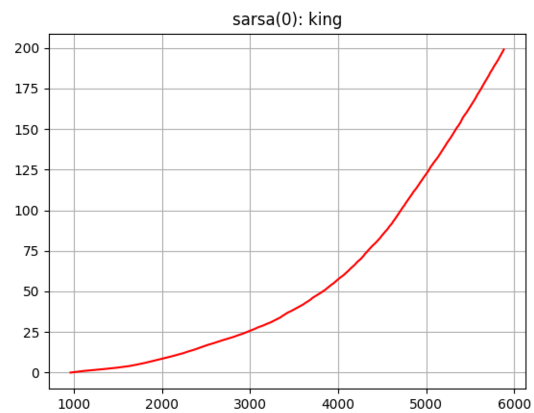


Figure 2: sarsa(0) on with king's actions

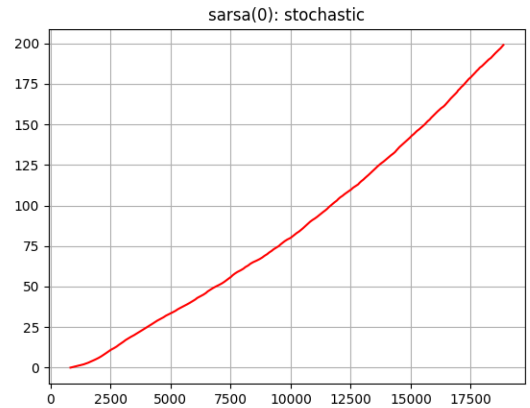


Figure 3: sarsa(0) on stochastic wind and King's moves

3 Comparative study

3.1 Comparison between updates

This section compares the performance of different update algorithms, namely, sarsa(0), expected sarsa, and Q-learning on the baseline model. The final cumulative steps needed for finishing 200 episodes are shown in Table 2. Figure 4 shows the plots of steps vs episodes. Q-learning seems to perform the best.

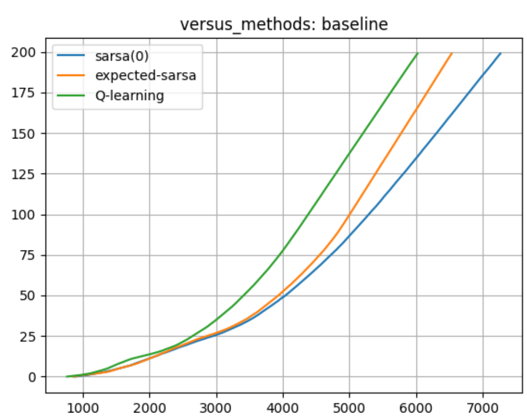


Figure 4: Different update rules on baseline model

Update	Steps
sarsa0	7264.4
Expected sarsa	6533.4
Q-learning	**6021.7**

Table 2: Steps after 200 episodes

3.2 Comparison between worlds

Figures 1, 2 and 3 have been summarised here. The final cumulative steps needed for finishing 200 episodes are shown in Table 3. Figure 5 shows the plots of steps vs episodes. The extra actions provided in King's world gives better performance than baseline model. When the wind becomes stochastic, it becomes more difficult to find the optimum Q-table. In fact, it does not seem to converge at all as the slope remains same throughout the 200 episodes.

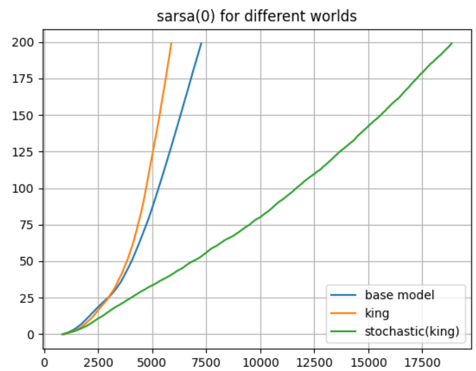


Figure 5: Performance of sarsa(0) on different worlds

World	Steps
Baseline	7264.4
King	5880.5
Stochastic	18588

Table 3: steps after 200 episodes

3.3 Comparing hyperparameters on the baseline model

This section compares the performance of Q-learning on setting different values of learning rates and ϵ. Figure 6 shows the effect of different epsilons for Q-Learning. We see that the best performance is got by setting epsilon equal to 0, which is quite interesting. The algorithm seems to converge for windy gridworld even when the policy is not GLIE.

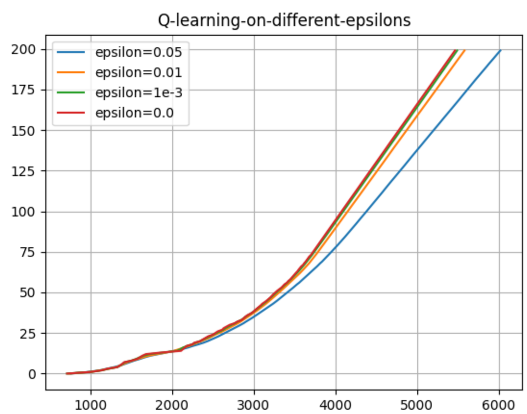
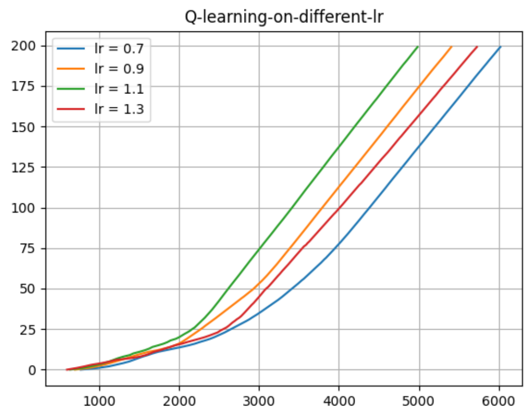


Figure 6: Performance of Q-learning with different epsilons

ϵ	Steps
0.05	6021.8
0.01	5583.14
$1e^{-3}$	5492.4
0.0	**5466**

Table 4: steps for different epsilons

Figure 7 shows the effect of different learning rates. We can observe a well expected 'minima' in the number of steps with regards to the learning rate(at a learning rate of 1.1 for Q-learning).



Learning rate	Steps
0.7	6021.72
0.9	5407
1.1	**4985.2**
1.3	5728.07

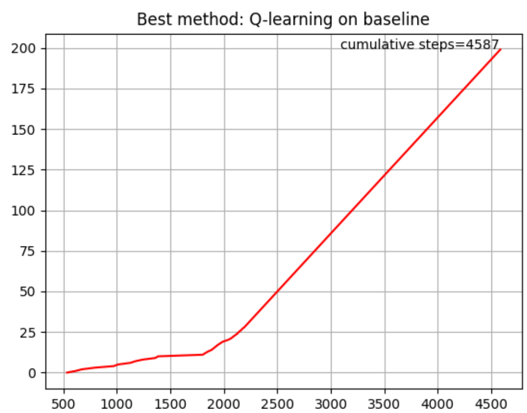
Table 5: steps for different learning rates

Figure 7: Performance of Q-learning with different learning rates

4 Best method

This section shows the best method(which converges the fastest) that could be obtained by tuning the hyperparameters and choosing between the three mentioned update methods for an agent in baseline world and in King's world.

4.1 Baseline model



World	Steps
ϵ	0.0
learning rate	1.1

Table 6: parameters for best method(baseline)

Figure 8: Best method on baseline: Q learning

4.2 King's model

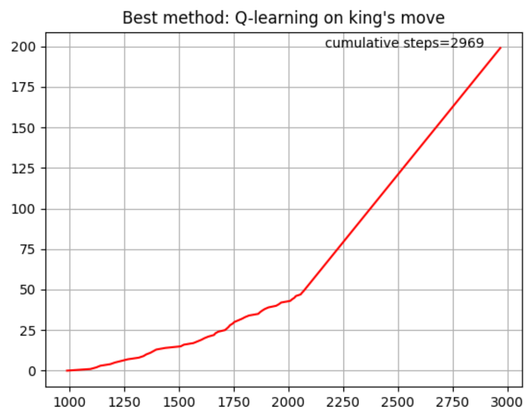


Figure 9: Best method on king: Q learning

World	Steps
ϵ	0.0
learning rate	1.2

Table 7: parameters for best method(king)