Project 3: Recreating Correlated Q-Learning

Myles Lefkovitz
gth836x
GitHub Hash: fad406afc7be6c77926aa70cfab7a990df613ac4

1. INTRODUCTION

In this project, I replicate Figure 3 from "Correlated Q-Learning" by Amy Greenwald and Keith Hall. Figure 3 includes 4 charts (a, b, c, and d) each demonstrating the Q-value error at each iteration of a soccer simulation.

2. SOCCER GAME

The grid:

The problem being solved is related to a small grid-based game with exactly 2 players called soccer. In this version of the problem, the grid has 8 spaces. Each player occupies exactly one space and the players cannot share a space. Four of the spaces are reserved for the goals (2 spaces each). See the image below (from the paper):

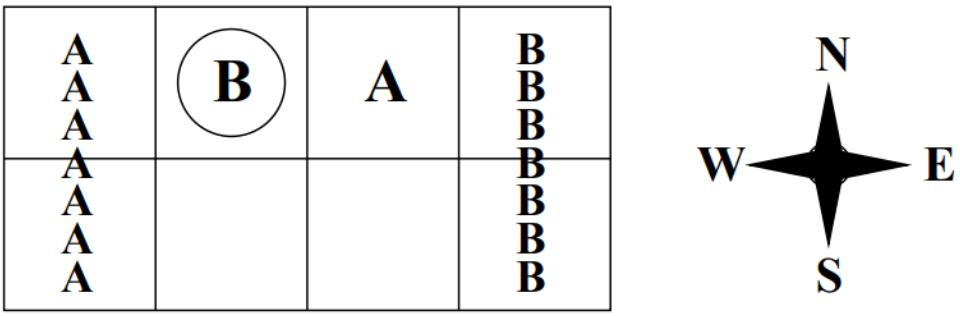


Figure 1. Soccer Grid (Greenwald 2003)

Scoring:

There is a single soccer ball. The game ends when the soccer ball reaches either goal. If the ball enters player A's goal (the one marked with As in the image) then A earns a +100 reward and B earns a -100 reward. If the ball enters player B's goal (the one marked with Bs in the image) then B earns a +100 reward and A earns a -100 reward.

The players:

Players A and B both select one of 5 possible actions (move N, E, S, W or 'stick'). The players' actions are executed in random order. Sometimes A goes first, sometimes B goes first. If the two players collide (want to reach the same space), only the first player moves. After a collision, the player that moved first ends up with the ball.

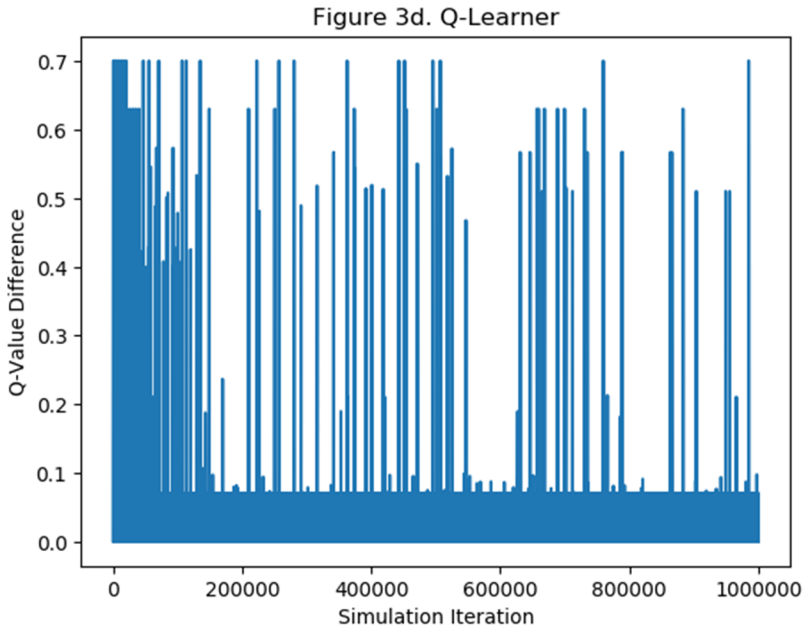
Simulation Implementation:

I implemented Pedro Velez's (github: bdvelez) environment for the soccer grid simulation (Velez, 2018).

3. EXPERIMENT 1: Q-LEARNING

In my first experiment I replicated the authors' implementation of a Q-learner on the soccer grid. I ran the Q-learner for 1,000,000 iterations, resetting the environment each time a reward was reached.

Below, find the error graph for this learner, corresponding to Figure 3d in the original paper.

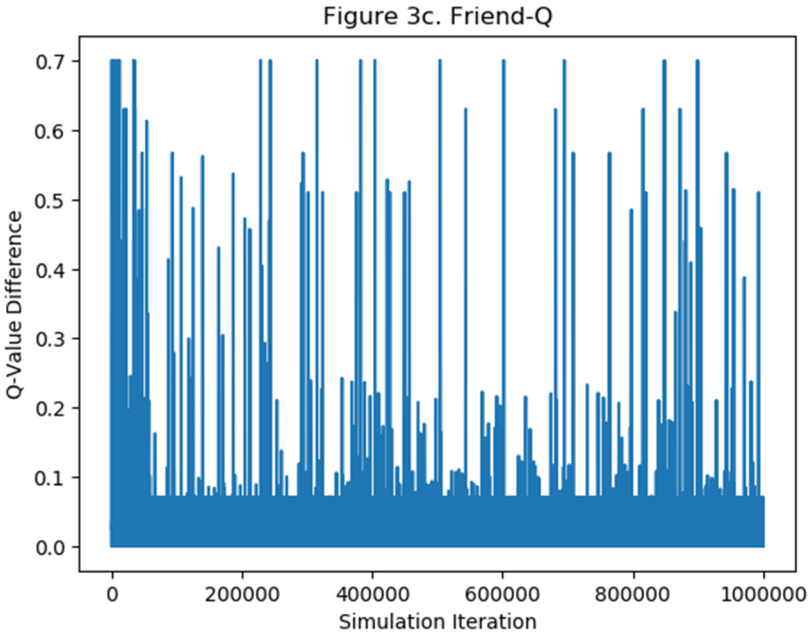


The graph is relatively similar to that in the paper (the graph in the paper has its y-axis cut at 0.5). Both this graph and that in the paper seem very jagged.

4. EXPERIMENT 2: FRIEND-Q

Using the experiment above (Q-learning) as a baseline, it was relatively straightforward to set up Friend-Q. As in the experiment above, I ran the Q-learner for 1,000,000 iterations, resetting the environment each time a reward was reached.

Below, find the error graph for this learner, corresponding to Figure 3c in the original paper.

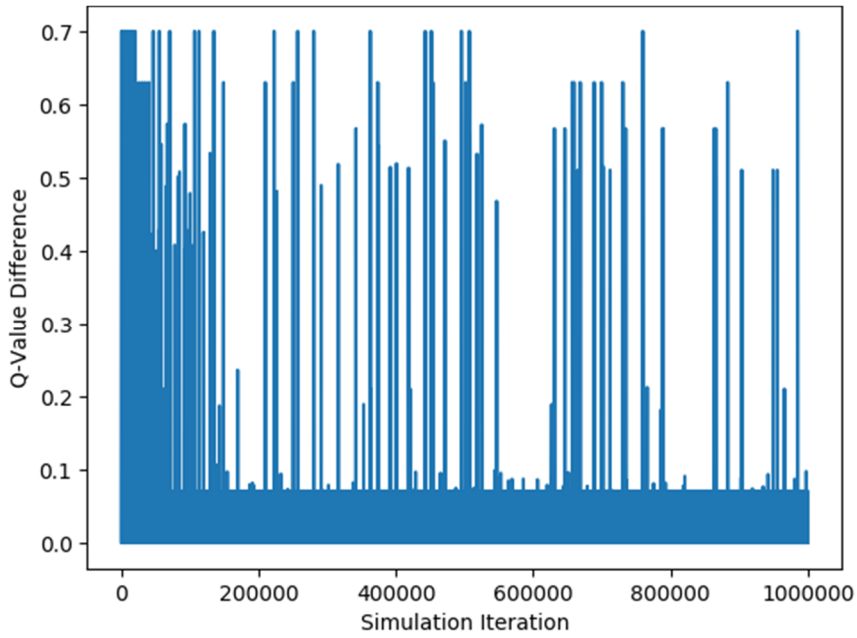


The graph is fairly dissimilar to that in the paper. The graph above appears far more jagged – it's not clear that this chart is converging, unlike that in the paper.

5. EXPERIMENT 3: FOE-Q

Using the experiments above (Q-learning and Friend-Q) as a baseline, I was able to set up Foe-Q. As in the experiments above, I ran the Q-learner for 1,000,000 iterations, resetting the environment each time a reward was reached.

Below, find the error graph for this learner, corresponding to Figure 3b in the original paper.



I'm not confident that I correctly implemented minimax since I avoided using an LP solver.

6. EXPERIMENT 4: CORRELATED-Q

Setting up Correlated-Q was very challenging given the variety of algorithms. I was unable to implement a working solution to this problem.

7. PROBLEMS ENCOUNTERED

What worked best

Nothing about solving this problem was easy or worked right out of the box. It was challenging to implement each algorithm. Further, it was challenging to understand the specifics that the papers (both the original and the more technical version) left out.

Graph Creation:

I used matplotlib and pyplot to create each graph automatically.

What didn't work

Development effort:

I wasn't able to figure out how to correctly implement the various correlated Q-learning algorithms. It was unclear to me which algorithm was used to create the error chart displayed in Figure 3a.

REFERENCES

1. Greenwald, A., Hall, K., & Serrano, R. (2003, August). Correlated Q-learning. In *ICML* (Vol. 3, pp. 242-249).
2. Greenwald, A., Hall, K., & Zinkevich, M. (2005). Correlated Q-Learning.
3. Littman, M. L. (2001, February). Friend-or-foe Q-learning in general-sum games. In *ICML* (Vol. 1, pp. 322-328).
4. Velez, P (2018). ML_Soccer. https://github.com/pdvelez/ml_soccer.