Udacity Robotics Nanodegree Term 2

DeepRL Project

Robert Aleck | github.com/mnbf9rca |16th August 2019

# Introduction

A Deep Q-Learning Network was implemented against a framework for a robotic arm with 3 DOF simulated within the Gazebo environment. The author had to:

* Subscribe to the relevant topics (camera imagery, collisions)
* Instantiate a DQN Agent
* Choose between velocity or position based arm joint manipulation, and implement in code
* Design an appropriate reward function, covering successful (collision between defined objects) and unsuccessful actions, including episode terminal states and interim rewards
* Tune hyperparameters and reward function for two scenarios:
  + At least 90% collision accuracy between any part of the arm and the collision object over at least 100 episodes, and
  + At least 80% accuracy in collisions between the gripper base of the robot arm and the collision object over at least 100 episodes

Reward functions

The robot arm control was selected as “position”, using default values.

At first, a large (>100 or even >500) REWARD\_WIN and REWARD\_LOSS value was used, with smaller/fractional rewards issued for interim states, however, the model failed to train effectively. Smaller values were then chosen, with additional multipliers providing greater weighting to episode-terminal (collision) events, as follows:

#define REWARD\_WIN 0.15f

#define REWARD\_LOSS -0.15f

#define REWARD\_COLISSION\_GROUND 10               // multiplier - it the ground

#define REWARD\_COLLISION\_CORRECT\_PART 20         // multiplier - hit the correct item

#define REWARD\_COLLISION\_WRONG\_PART 10           // multiplier - hit the wrong item

#define MIN\_DISTANCE\_TO\_MOVE\_WITHOUT\_PENALTY 0.05f // how far must the gripper have moved compared to last frame to avoid a penalty

The reward function was unchanged when training to hit the gripper base. The selection of collision target is changed with the REWARD\_ANY\_COLLISION parameter – when set to false, only collisions between COLLISION\_ITEM and COLLISION\_POINT were rewarded, otherwise all collisions with COLLISION\_ITEM resulted in a reward.

#define REWARD\_ANY\_COLLISION false               // reward for hitting any part of the arm on the tube

The following rewards were issued for episode-terminal states:

|  |  |
| --- | --- |
| Ground contact | REWARD\_LOSS \* REWARD\_COLISSION\_GROUND |
| Collision between COLLISION\_ITEM and either any other part, or with COLLISION\_POINT | REWARD\_WIN \* REWARD\_COLLISION\_CORRECT\_PART |
| Collision between COLLISION\_ITEM and any part other than COLLISION\_POINT | REWARD\_LOSS \* REWARD\_COLLISION\_WRONG\_PART |

To encourage the arm to move towards the target, interim rewards were issued. A positive movement towards the goal received REWARD\_WIN \* timePenalty (where timePenalty is a value trending from 1 to 0 proportional to the number of frames completed), whereas movement away received REWARD\_LOSS. Additionally, as it was noticed that the arm spent a long time in a relatively stable position, REWARD\_LOSS is added to rewardHistory at each frame that the gripper failed to move, in a smoothed, moving average, at least MIN\_DISTANCE\_TO\_MOVE\_WITHOUT\_PENALTY since the previous frame.

Hyperparameters

The following hyperparameters were chosen and used for both collision models:

#define INPUT\_WIDTH 64

#define INPUT\_HEIGHT 64

#define OPTIMIZER "Adam"

#define LEARNING\_RATE 0.1f

#define REPLAY\_MEMORY 10000

#define BATCH\_SIZE 512

#define USE\_LSTM true

#define LSTM\_SIZE 256

The input image size was observed to be 64x64, so this was selected for the model. Adam was selected for the optimizer based on previous experience with reinforcement learning, where the author has found it to converge quicker for relatively simple models. The learning rate was initially 0.01, but was increased to 0.1 to speed up learning – no evidence of overfitting was found. Replay memory was left as default. Batch size was increased to take advantage of the large memory available on the server, and the LSTM size increased experimentally – higher than 256 failed to fit, lower than 128 slowed learning

Results

Results: Explain the results obtained for both objectives. Include discussion on the DQN agent's performance for both objectives. Include watermarked images, or videos of your results.

Student should describe and briefly explain the results they achieved for both objectives. The discussion should also include their comments on the DQN agent's performance and if there were any shortcomings. Student should include either watermarked images of their results, or attach a video that displays the results and the arm in action.

# References

|  |  |
| --- | --- |
| [1] | M. Labbé, “RTAB Map - Real-Time Appearance-Based Mapping,” 2017. [Online]. Available: http://introlab.github.io/rtabmap/. [Accessed 21 07 2017]. |
| [2] | r. aleck, “github/mnbf9rca/robond\_term2\_p2\_where\_am\_i,” 2019. [Online]. Available: https://github.com/mnbf9rca/robond\_term2\_p2\_where\_am\_i. |
| [3] | M. Montemerlo, S. Thrun, D. Koller and B. Wegbreit, “FastSLAM: A Factored Solution to the Simultaneous Localization and Mapping Problem,” *Proceedings of the AAAI National Conference on Artificial Intelligence.,* pp. 593-598, 2002. |
| [4] | S. Thrun, Y. Liu, D. Koller, A. Y. Ng, Z. Ghahramani and H. Durrant-Whyte, “Simultaneous localization and mapping with sparse extended information filters,” *The International Journal of Robotics Research,* vol. 23, no. 7-8, 2004. |
| [5] | S. Thrun, W. Burgard and D. Fox, “11: THE EXTENDED INFORMATION FORM ALGORITHM,” in *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*, MIT Press, 2005, pp. 267-268. |
| [6] | M. Montemerlo, S. Thrun, D. Koller and B. Wegbreit, “FastSLAM: A Factored Solution to the Simultaneous Localization and Mapping Problem,” *In Proceedings of the AAAI National Conference on Artificial Intelligence,* Vols. 593-598, 2002. |
| [7] | G. Grisetti, R. Kummerle, C. Stachniss and W. Burgard, “A Tutorial on Graph-Based SLAM,” *IEEE Transactions on Intelligent Transportation Systems Magazine,* vol. 2, no. December, pp. 31-43, 2010. |