CaRL:
Learning Lane-Sharing using a Deep Q-Network

Zoë D. Papakipos
CSCI1970: Independent Research

I. ABSTRACT

This semester in the RLab Jake Beck, Matt Cooper, Michael Gillett, JD Fishman, and I researched deep reinforcement learning for self-driving cars under the guidance of Professor Michael Littman. We spent the first part of the semester setting up our Unity simulator (which is built on top of Udacity's) and networking between the simulator and Python scripts so we could run deep learning algorithms in Tensorflow in conjunction with the simulator. We then moved on to a more socially complex problem: lane-sharing with oncoming traffic when part of the road is blocked.

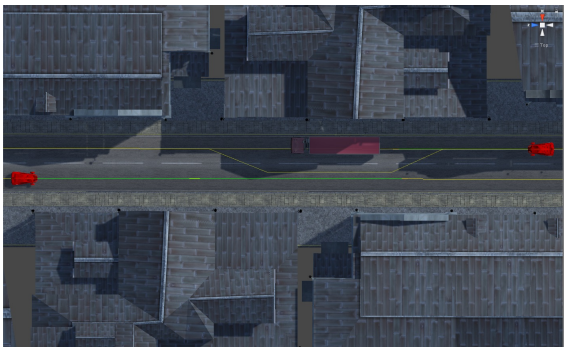

Fig. 1. The task set up in Unity

We built a DQN, or Deep Q-Network, a neural net architecture introduced by DeepMind [1], to control CaRL's actions. Like DeepMind's, our DQN uses a neural net to predict the Q-value, or expected discounted future reward, of a state CaRL is in considering different next actions. The DQN uses unsupervised learning but trains the Q-net in a supervised way by using the Temporal Difference Error to estimate the loss of each prediction. We give CaRL varying amounts of reward after every action to encourage reaching the end of the street quickly, but also discourage crashing. We ran approximately 2.5 million training batches, randomly sampled from an experience buffer of the past $50,000$ time steps.

We trained CaRL against a simple AI car. This simple AI will let CaRL pass $\frac{2}{3}$ of the time, but otherwise will not yield. CaRL therefore has to learn caution and treat the other car as somewhat unpredictable. We plotted CaRL's reward per episode when training against the AI car. We also plotted the average reward we, the researchers, were able to get controlling CaRL, as well as the average reward an agent acting randomly was able to get over the same number of episodes. These baselines act as a rough upper and lower bound for how good we can expect CaRL to do. We smoothed the data such that each plotted data point is the average of the last 50 episodes, rather than the raw reward of each episode.

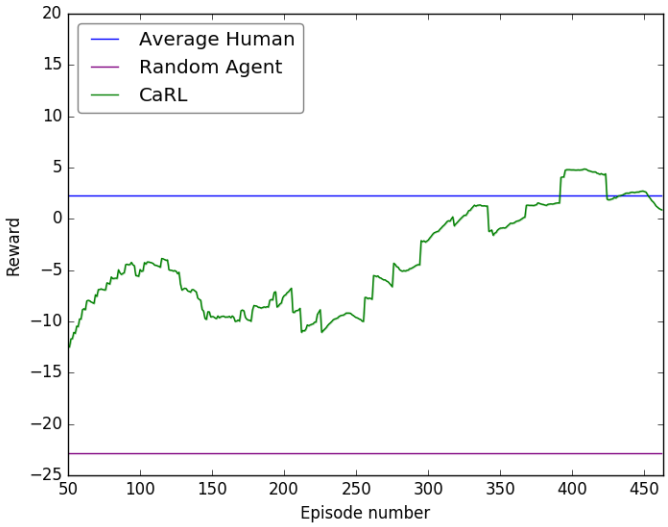

Fig. 2. CaRL learns to average even more reward than we (the researchers) did, and far better than an agent acting randomly.

As you can see, CaRL performs significantly better than the random agent in all episodes, meaning that it starts learning pretty quickly. CaRL approaches and even sometimes surpasses the average human value toward the end of training. The average human value reflects a policy of generally waiting to see if the dumb AI will let us pass, and then gunning it the rest of the way. However, we humans sometimes deviated from this optimal behavior because we thought we could predict what the other car would do, and perhaps because humans are inherently risky. CaRL, and AI models in general, lack human impulsiveness and thus have the ability to avoid danger much better than people. With the proper algorithms, self-driving cars have the potential to save countless lives and greatly improve road safety.

II. REFERENCES

[1] Mnih et al. Human-level control through deep reinforcement learning. Nature, 2015. URL: https://storage.googleapis.com/deepmind-media/dqn/DQNNaturePaper.pdf