CS440: Reward Function
Drew Kristensen

Reward

The reward is an important part of reinforcement learning process since it is the metric that the Q-function tries to approximate and is thus what the network works to maximize. As mentioned in Genders and Razavi [1], there are multiple rewards to choose from in this scenario: the change in the time vehicles are waiting, the change in the number of vehicles stopped, and the change in total vehicle throughput. Genders and Razavi used the change in cumulative vehicle delay between actions, which proved to be successful in their application to the problem. While this is a good idea, the reward specified by this research will be the cumulative squared vehicle delay. That is, the sum of the delays squared for each individual vehicle between each action. The thought behind it is this: with 1 car, they would spend a time linear to each car on the orthoganal road, and given enough traffic on the orthoganal road, they have the ability to be stuck for a long period of time. While this doesn't lead to the most efficient solution, this is a solution that I see being more popular with drivers. This however can change, as the only difference between linear and squared factors of this metric are the power we raise the values to.

References

[1]   Wade Genders and Saiedeh Razavi. "Using a Deep Reinforcement Learning Agent for Traffic Signal Control". In: *CoRR* abs/1611.01142 (2016). arXiv: 1611.01142. URL: http://arxiv.org/abs/1611.01142.