Efficient policy learning from rare states leading to Queue Blocking

Daniel Mastropietro

March 18, 2020

1 Introduction

Following the paper by Massaro et al. [1], we propose a reinforcement learning algorithm that learns the optimal policy of an M/D/2/1 queue with the aim of minimizing the probability of blocking.

The focus of the designed algorithm is to learn the optimal policy in an efficient way, where efficiency is thought of in terms of learning from the state-action pairs that are most informative of rewards.

The algorithm leverages the knowledge we have in advance about rewards in the system. In fact, the reward received by the agent is always zero except at the state-action pairs that lead to blocking, in which it receives a negative reward.

The problem can be formally stated as follows: one single queue of capacity K receives requests to process two types of jobs, each with a Poisson-like arrival rate λ_i and each being served in deterministic time, assumed to be equal for all job types. The agent needs to define a policy that decides whether to accept or not a new job of a given class arriving to the queue.

2 Markov Decision Process model

We model the queue and the managing agent with a Markov Decision Process (MDP) having the following characteristics:

States: Occupancy level of the queue $s \in \{0, ..., K\}$, i.e. there are $K+1$ possible states.

Note that the state only records the *total* number of jobs in the queue. This limits policies to functions of the total queue occupation, disrespectful of their type.

Actions: $a \in \{0, 1\}$ (i.e. reject or accept a new job, respectively). Both actions are possible for all states except for the last one $s = K$ for which only $a = 0$ is possible.

Rewards: the system gives a large penalty when the queue blocks. For all other states, the reward is set to 0 for all actions.

The transition probabilities are known: when a new job is accepted, the state always increases by 1.

3 Prediction problem

For the prediction problem, we set the policy to be fixed and to always accept an incoming job.

4 Control problem

References

[1] Antonio Massaro, Francesco De Pellegrini, and Lorenzo Maggi. Optimal trunk-reservation by policy learning. In *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*. IEEE, apr 2019.