Q-learning for the Optimal Bernoulli Routing

Model Description

We consider a system with N parallel queues and a single dispatcher (or router). Jobs arrive to the system according to a Poisson process of rate λ. We assume that the service time of jobs in the queues is exponentially distributed and we denote by r_i the rate at which jobs at queue i are served. When a job arrives to the dispatcher it is immediately routed to Queue i with probability p_i. Hence, $\sum_{i=1}^{N} p_i = 1$. We aim to study the value of the routing probabilities such that the mean number of customers is minimized.

The authors in [1] study this system and characterize the optimal routing probability. Here, aim to show that Q-learning can be used to learn which is the optimal routing probability.

Markov Decision Process Formulation

We formulate the above problem as a Markov Decision Process in discrete time in which the discretization is carried out when a job arrives to the system. We consider the discounted cost problem, that is, we aim to find the probabilities $p_1, p_2, \ldots, p_N$ such that the following expression is minimized:

$$\mathbb{E}\left[\sum_{t=0}^{\infty}\sum_{i=1}^{N}\delta^t Q_i(t))\right], \tag{1}$$

where $\delta \in (0,1)$ and $Q_i(t)$ is the number of jobs in Queue i at time slot t.

Let us define the following elements of the Markov Decision Process we consider:

- The state represents the number of jobs in each queue. Therefore, the set of states is a vector of size N such that each element is 0 or a natural number. That is, $\mathcal{S} = \mathbb{N}_0 \times \cdots \times \mathbb{N}_0$, where $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$.

- The action is a vector $(p_1, \ldots, p_N)$ such that the i-th element is the probability that an incoming job is sent to Queue i. Therefore, the set of actions is a probability vector of size N. We assume that each element of the probability vector belongs to $\{0, \frac{1}{d}, \frac{2}{d}, \ldots, \frac{d-1}{d}, 1, \}$ for a fixed d (this means that the probabilities will never be real values).

- The cost is the total number of customers in the system when a job is sent to one of the queues.

- The transition probabilities. Between two arrivals, one or more jobs can be served at Queue i. We denote by $q_{i,j}$ the probability that j jobs are served at Queue i in a interval of time of λ.

References

[1] E. Altman, U. Ayesta, and B. Prabhu. Load Balancing in Processor Sharing Systems. *Telecommunication Systems*, 47(1-2):pp.35–48, May 2011.