Automatic Goal Generation for Reinforcement Learning Agents

Alessandro Trenta - 566072

ISPR - Dip. of Informatics
Università di Pisa

25 maggio 2022

The problem

- In general RL frameworks the main objective is to find a policy $\pi(a_t|s_t)$ that maximizes the expected future return.
- In this problem, instead of maximizing the return over a single reward function we want to analyze the situation in which we have a range of reward functions r^g indexed with a goal $g \in \mathcal{G}$.
- A goal is defined as a set of states $\mathcal{S}^g \subset \mathcal{S}$. The reward function associated to this goal is

$$r^g(s_t, a_t, s_{t+1}) = \mathbb{1}\{s_{t+1} \in \mathcal{S}^g\}$$

- We will consider $S^g = \{s \in \mathcal{S} : d(f(s), g) \leq \epsilon\}$ where f projects the states in the goal space $\mathcal{G}$ and d is a distance in this space.

- Given g we consider a Markov Decision Process that terminates whenever $s_t \in \mathcal{S}^g$. We then consider the return to be $R^g = \sum_{t=0}^{T} r_t^g$. This is actually a binary random variable and indicates whether the agent is able to reach a state close enough to the goal in at maximum T time steps.
- The policy is also g dependent: $\pi(a_t|s_t, g)$ and the expected return for a goal is

$$\begin{aligned} R^g(\pi) &= \mathbb{E}_{\pi(\cdot|s_t,g)} \left[\mathbb{1}\{\exists t \in [1, \dots, T] : s_t \in \mathcal{S}^g\right] \\ &= \mathbb{P}\left(\exists t \in [1, \dots, T] : s_t \in \mathcal{S}^g\right) \end{aligned}$$

- We then assume to have a test distribution over goals p_g so that out objective is to maximize the expected mean return over goals obtaining the policy

$$\pi^*(a_t|s_t, g) = \arg\max_{\pi} \mathbb{E}_{g \sim p_g(\cdot)}[R^g(\pi)]$$

which is indeed the average probability of success over all possible goals (w.r.t. p_g).

Three main assumptions:

- A policy learned from enought goals in specific area of $\mathcal{G}$ can learn to interpolate well for other goals in this area.
- A policy trained on some set of goals is a good initialization for other goals that are close enough.
- If a goal is reachable, there is a policy that can reach it consistently.

Bibliografia I

 Carlos Florensa, David Held, Xinyang Geng, and Pieter Abbeel. Automatic goal generation for reinforcement learning agents, 2017.

Grazie per l'attenzione