Mapless Navigation with Deep Reinforcement Learning

Apotheoses and Momentousness

- **AlphaGo** → AlphaGo Zero → AlphaZero → MuZero
- **ChatGPT** fine-tuned in a process called *reinforcement learning from human feedback*
- Ability to tackle problems that are too complicated to model accurately with traditional approaches

Aspirations

- **Implement** SOTA DRL algorithms
- **Implement** training environments in a simulator
- **Train** policies and **evaluate** them

Attainments

- **Implemented** a library and a simulation environment
- Large amount of **experiments** were conducted
- **Trained** polices in a small empty indoor environment
- Developed **skills** and **insights** on DRL and AI

Design for the DRL Library

- **Modular** and **composable**

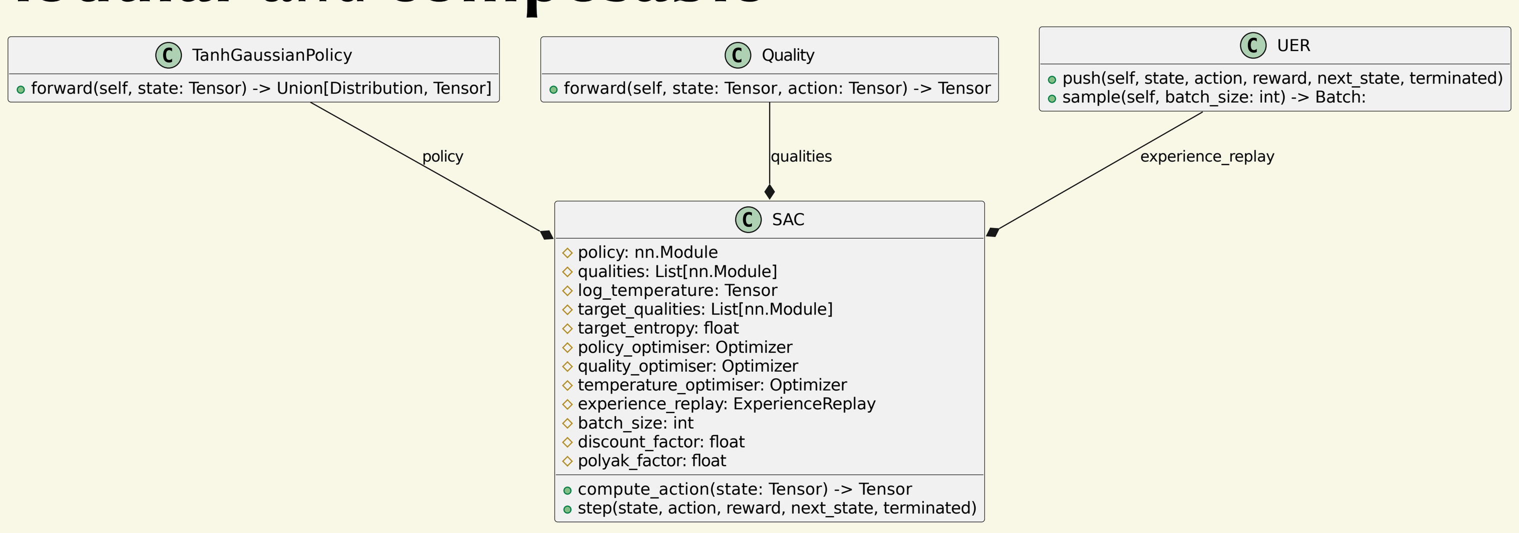

Figure: UML class diagram of the Soft Actor-Critic algorithm

- **Unit-tested** with *pytest* and **integration-tested** on various *Gymnasium's* built-in environments
- **Packaged** with the latest PyPA specification

Soft Actor-Critic

Optimal policy $\pi^* = \arg\max_\pi \sum_t \mathbb{E}_{(s_t, a_t)\sim\rho_\pi} [r(s_t, a_t) + \alpha\mathcal{H}(\pi(\cdot|s_t))]$
Initialise critic networks Q_{θ_1}, Q_{θ_2}, and actor network π_ϕ with random parameters θ_1, θ_2, ϕ
Initialise target network weights $\overline{\theta}_i \leftarrow \theta_i$ for $i \in \{1, 2\}$
Initialise an empty replay buffer $\mathcal{B}$
for each iteration **do**
  for each environment step **do**
    $a_t \sim \pi_\phi(\cdot|s_t)$
    $s_{t+1} \sim p(\cdot|s_t, a_t)$
    $\mathcal{B} \leftarrow \mathcal{B} \cup \{(s_t, a_t, r(s_t, a_t), s_{t+1})\}$
  for each gradient step **do**
    $\theta_i \leftarrow \theta_i - \lambda_Q \hat{\nabla}_{\theta_i} \mathcal{J}_Q(\theta_i)$ for $i \in \{1, 2\}$
    $\phi \leftarrow \phi - \lambda_\pi \hat{\nabla}_\phi \mathcal{J}_\pi(\phi)$
    $\alpha \leftarrow \alpha - \lambda \hat{\nabla}_\alpha \mathcal{J}(\alpha)$
    $\overline{\theta}_i \leftarrow \tau\theta_i + (1-\tau)\overline{\theta}_i$ for $i \in \{1, 2\}$

where $\hat{\nabla}\mathcal{J}$ denotes the estimated gradient of an objective function, $\overline{\theta}$ denotes an exponentially moving average of θ, λ denotes the learning rate, and α is the temperature parameter that determines the relative importance of the entropy term versus the reward.

Design for the Training Environment

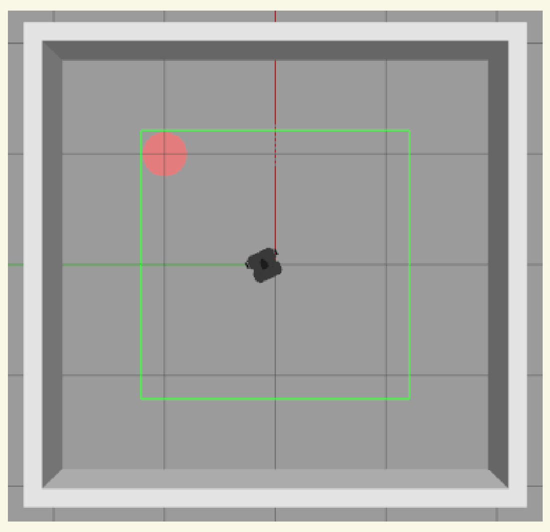

Figure: Aerial view of the venue in Gazebo Classic

- Observed state space:
$$\mathcal{S} = (x, d),$$
where x is the data from LiDAR, and d is the relative distance of the navigation goal w.r.t the robot.
- Action space:
$$\mathcal{A} = (v, \omega),$$
where v and ω are respectively the linear and angular velocity of the robot.
- Reward shaping
$$\mathcal{R}(s_t, a_t) = \begin{cases} \mathcal{R}_{\text{goal}}, & \text{if } d_t < d_{\text{th}} \\ \mathcal{R}_{\text{obstacle}}, & \text{if } x_t < x_{\text{th}} \\ c(d_t - d_{t-1}), & \text{otherwise.} \end{cases}$$
Notice that c is an amplification factor and a parameter of the environment, $s_t \in \mathcal{S}$, and $a_t \in \mathcal{A}$.

Limitations of Contemporary DRL

- **Partial** observability
- **Sparse** rewards
- Catastrophic forgetting

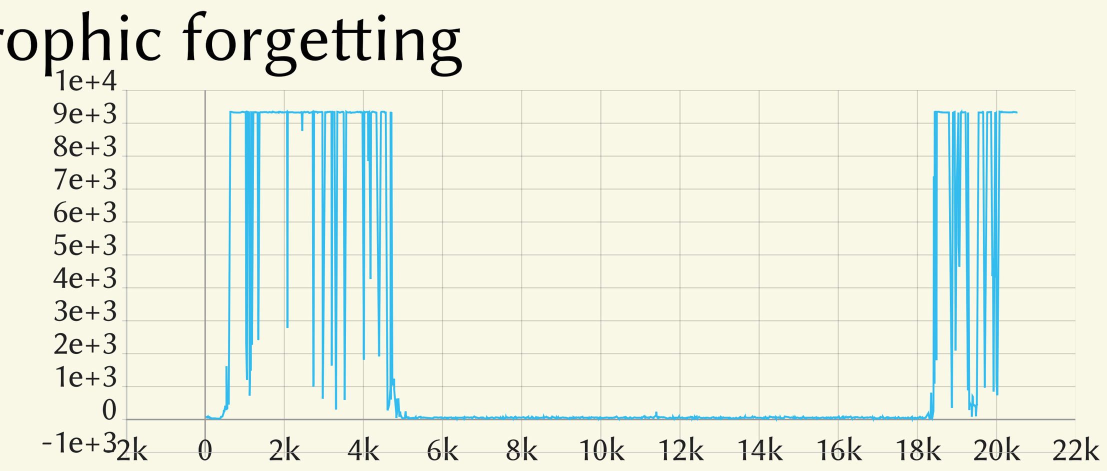

Figure: Episodic return of TD3 on InvertedDoublePendulum

- Transfer learning
- Exploration-exploitation trade-off

Conclusions and Future Work

- Explored DRL by DIY and experimenting
- This case study demonstrates benefits and shortcomings of DRL
- Upgrade to non-stationary env to explore what difference would make on the policies trained

Selected References

- L. Tai, G. Paolo, and M. Liu, "Virtual-to-real deep reinforcement learning: Continuous control of mobile robots for mapless navigation," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 31-36.
- T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *Int. Conf. on Machine Learning*, 2018.
- P. Henderson, R. Islam, P. Bachman, J. Pineau, D. Precup, and D. Meger, "Deep reinforcement learning that matters," in *AAAI Conf. on Artificial Intelligence*, 2018.

Xu, Minghong (Student ID: 201601082)
Supervisor: Dr Murat Üney | Assessor: Prof. Jason Ralph
Department of Electrical Engineering and Electronics, University of Liverpool