УДК 004.896

Multi-task oriented deep reinforcement learning for robot navigation

Бархум М (ИТМО), **Динь Нгок Туан** (ИТМО),
Научный руководитель – профессор, Пыркин А.А.
(ИТМО), **Консультант, - Борисов О.И**(ИТМО),

Introduction. Navigating mobile robots through complex and dynamic environments poses a significant challenge in the field of robotics. While Deep Reinforcement Learning (DRL) has advanced robotic control, persistent issues such as suboptimal policies, reward function design complexity, and the inability of Deep Neural Networks (DNNs) to generalize across competing objectives (e.g., speed vs. safety) limit performance. Manual reward shaping, which requires balancing penalties and incentives for diverse tasks, often leads to unintended behaviors or oversimplified policies. In this work, we address these limitations by proposing a multi-task reinforcement learning framework for a ground mobile robot, enabling simultaneous optimization of goal-oriented navigation and obstacle avoidance through a structured reward decomposition.

Main part. The Twin Delayed Deep Deterministic Policy Gradient (TD3) algorithm is uniquely suited for navigation tasks due to its robustness against overestimation bias—a common pitfall in value-based RL methods. TD3 achieves this through twin Q-networks, which independently estimate state-action values and use the minimum of both estimates for policy updates, thereby curbing excessively optimistic value approximations. Additionally, delayed policy updates decouple actor and critic training frequencies, stabilizing learning by reducing variance in the target values. To further enhance TD3 for multi-objective navigation, we introduce task-specific multi-head networks, each dedicated to a distinct sub-goal:

1) Goal-reaching head: Optimizes for time-efficient navigation using a reward function that incentivizes path shortening and penalizes excessive traversal time.
2) Collision-avoidance head: Focuses on obstacle-free path planning by assigning penalties for proximity to obstacles and collisions

By disentangling reward signals, the agent learns decoupled policies for each sub-task, which are then synthesized into a unified navigation strategy. This architecture minimizes conflicting gradient updates during training, a common issue in monolithic reward structures, while preserving the sample efficiency and stability of TD3.

Conclusion. The framework is validated in a ROS 2.0 Gazebo simulation, where a differential-drive mobile robot operates in a dynamic environment with static and randomized obstacles. Goals are randomized at each episode to test generalization. Quantitative results demonstrate the robot’s ability to (1) maintain a 95% collision-free success rate in cluttered settings, and (2) adapt to unseen environments by leveraging learned priors from multi-task training. These outcomes highlight the system’s potential for real-world deployment, where balancing efficiency and safety is paramount. Future work will extend this approach to multi-agent scenarios and real-time hardware validation.

References.

1. Cimurs, R., Suh, I.H. and Lee, J.H., 2021. Goal-driven autonomous exploration through deep reinforcement learning. *IEEE Robotics and Automation Letters*, *7*(2), pp.730-737.

2. Bai, S., Song, S., Liang, S., Wang, J., Li, B. and Neretin, E., 2022. UAV maneuvering decision-making algorithm based on twin delayed deep deterministic policy gradient algorithm. *Journal of Artificial Intelligence and Technology*, *2*(1), pp.16-22.

3. S. Sodhani, A. Zhang, J. Pineau, Multi-task reinforcement learning with context-based representations (2021). arXiv:2102.06177. URL https://arxiv.org/abs/2102.06177

Бархум М (автор) Подпись

Пыркин А.А (научный руководитель) Подпись

Динь Нгок Туан (соавтор) Подпись

Борисов О.И (Консультант) Подпись