GLOBAL RESEARCH IMMERSION PROGRAM FOR YOUNG SCIENTISTS

Reproducing Paper: Adaptive Planner Parameter Learning from Reinforcement

Zecheng Zhu
Department of Electronic and Electrical Engineering, Imperial College London;
Hosted at Zhejiang University (ZJU)

Introduction

In the rapidly evolving field of robotic navigation, the ability of autonomous systems to adapt to new environments without extensive human intervention remains a paramount challenge. Traditionally, navigation systems rely on a fixed set of parameters optimized for specific operational conditions. These parameters, such as maximum speed, sampling rate, and inflation radius, often require expert tuning to adapt to different environments, making the deployment process both time-consuming and resource-intensive. Moreover, methods that learn from human demonstrations, while reducing the need for expert tuning, are generally limited by the demonstrators' performance and the specific conditions of the environment.

Recognizing these limitations, the study of Adaptive Planner Parameter Learning from Reinforcement (APPLR) proposes an innovative approach, utilizing reinforcement learning (RL) to dynamically adjust navigation parameters in real time. This method offers the potential for autonomous systems to self-adapt to new environments, learning optimal navigation strategies through trial and error without reliance on predefined parameter sets or human input.

This poster presents a comprehensive reproduction of the APPLR study, originally conducted by researchers at the University of Texas at Austin. The objective of this reproduction is to validate the original findings and further explore the efficacy and adaptability of the APPLR approach in both simulated and real-world settings. By closely following the methodologies and experimental setups described in the original study, this work aims to assess the robustness of the APPLR framework and its potential applications in broader autonomous navigation contexts.

The following sections detail the research objectives, the methods replicated from the original study, and the results of this reproduction effort, highlighting both the challenges encountered and the insights gained. This reproduction not only serves to confirm the significant capabilities of APPLR but also contributes to the ongoing discourse on enhancing the autonomy and efficiency of robotic navigation systems.

Objective

• Validate the effectiveness of the Adaptive Planner Parameter Learning from Reinforcement (APPLR) method as initially presented. This includes reproducing the original study's approach to dynamically adjust navigation parameters through reinforcement learning, aiming to demonstrate that APPLR can autonomously adapt to new environments without expert tuning or human demonstration.

• Evaluate the generalizability of the APPLR framework across a variety of environments, both simulated and physical, as detailed in the original publication. This involves rigorously testing the reproducibility of the original results to understand the conditions under which APPLR excels or faces limitations.

• Analyze the robustness and efficiency of the reinforcement learning strategies employed by APPLR. This includes a detailed assessment of the learning algorithms, reward structures, and parameter policies to identify potential areas for further improvement or optimization.

By addressing these objectives, this reproduction study seeks to not only corroborate the findings of the original APPLR study but also to contribute to the broader scientific discussion regarding adaptive learning technologies in autonomous navigation systems.

Method

APPLR Approach

The APPLR (Adaptive Planner Parameter Learning from Reinforcement) model is a novel approach that enhances the adaptability of traditional navigation systems through the integration of reinforcement learning (RL). In reproducing this method, the focus is on how APPLR allows for dynamic adjustments of navigation parameters, which is in stark contrast to the static parameter sets typically used in classical navigation frameworks.

Integration with Classical Systems: The original APPLR framework integrates with classical motion planners by treating them as components of a broader, more complex meta-environment. In this reproduction, we maintain this integration by interfacing the RL model with the existing motion planning algorithms. This setup enables the RL agent to interact directly with both the motion planner and the real-world or simulated environment, providing a comprehensive framework for testing and validation.

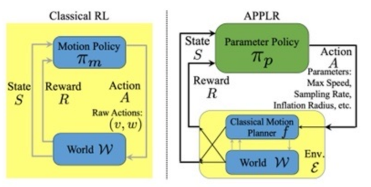


Fig. 1: Instead of learning an end-to-end motion policy πm which takes state S and reward R from the world W and produces raw actions A, e.g. linear and angular velocity (v, ω) (left), APPLR treats an underlying classical motion planner f as part of the meta-environment E (along with the world W) and the learned parameter policy πp interacts with it through actions in the parameter space (right). In this way, the RL agent selects its action in the form of a set of navigation parameters at each time step and reasons about potential future consequences of those parameters, rather than tuning a single set of parameters for the entire environment only considering the current situation.

Real-Time Optimization: APPLR's core functionality lies in its ability to optimize navigation parameters in real-time. The RL agent continuously evaluates the current state of the environment along with historical data to predict future conditions. In our reproduction, we simulate this by repeatedly exposing the system to a variety of environmental scenarios, measuring the system's adaptability and response accuracy under changing conditions. This aspect is crucial for demonstrating the practical utility of APPLR in real-world navigation tasks.

Parameter Policy

The parameter policy is the mechanism through which APPLR makes its dynamic adjustments. This policy is not fixed but is instead learned through continuous interaction with the environment, leveraging reinforcement learning to refine its decisions over time.

Dynamic Selection of Parameters: In our reproduction, the policy is trained to dynamically select appropriate navigation parameters at each decision point. This training process involves both on-policy and off-policy RL techniques to balance exploration of new strategies with exploitation of known successful parameters. We emulate the training conditions described in the original study to ensure that the learning outcomes are comparable.

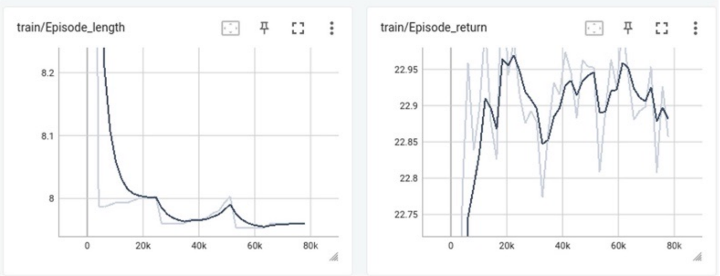
By meticulously following these methods and ensuring that each aspect of the APPLR approach is accurately replicated, this reproduction study aims to validate the original findings and potentially highlight areas for further improvement. The fidelity of the reproduction process is critical for confirming the versatility and effectiveness of APPLR as a pioneering solution in adaptive navigation systems.

Results & Conclusion

Results

Training Performance

The graphs below depict the training performance of the APPLR system during the reproduction study. Two key metrics are illustrated: Episode Length and Episode Return.

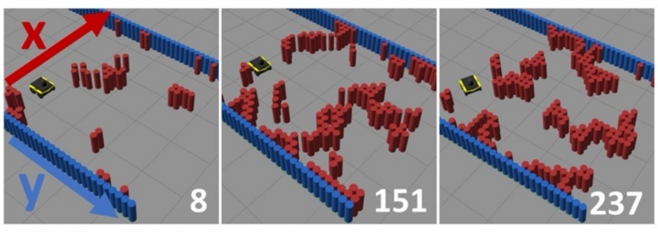


Episode Length: This metric shows a significant decrease over the course of training, indicating that the APPLR model is learning to reach its goal more efficiently. Initially, the episode length is variable, as the system explores different strategies. After approximately 40k training steps, there is a noticeable stabilization in episode length, suggesting that the APPLR has started to converge on effective navigation parameters.

Episode Return: Reflecting the cumulative reward per episode, this graph demonstrates fluctuations early in training, which is typical in reinforcement learning scenarios as the agent explores various actions. As training progresses, the returns show an upward trend, peaking around 60k steps before slightly tapering off. This pattern signifies the system's increasing proficiency in navigating the environment while maximizing the reward function components, particularly those related to reaching the goal efficiently and avoiding obstacles.

Environmental Interaction

The sequence of images below demonstrates the APPLR system interacting with a simulated environment at various difficulties for navigation of a training episode:



Low (World Index 8): Only a few obstacles.

Mid (World Index 151): More obstacles but easy to find feasible path.

High (World Index 237): More obstacles & hard to find feasible path.

Conclusions

This reproduction study confirms the effectiveness of the APPLR approach in enhancing the adaptability and efficiency of autonomous navigation systems. The training results corroborate the original findings that reinforcement learning can significantly improve the performance of navigation systems in dynamic environments.

Key Findings:

•**Adaptability**: The APPLR model demonstrated a strong ability to adapt to complex environments, as evidenced by the progressive decrease in episode lengths and improvement in episode returns.

•**Efficiency**: The system learned to optimize its navigation paths quickly, resulting in faster goal attainment and higher cumulative rewards.

•**Robustness**: Through continuous interaction with a range of simulated environments, APPLR proved robust, handling diverse and challenging navigation scenarios effectively.

Future Work

The successful reproduction of the APPLR study provides a strong foundation for further advancements in the field of autonomous navigation. A notable direction for future work involves addressing a critical issue identified during the study: the potential discrepancy between the planned actions by the navigation system and the actual movements executed by the robot. This discrepancy can arise due to various factors, including mechanical limitations, environmental variables, and sensor inaccuracies.

Development of an Adaptive Controller Using Reinforcement Learning

To mitigate the aforementioned issue, the next phase of research will focus on developing an adaptive controller using reinforcement learning (RL). This controller will aim to ensure that the robot's physical movements align more closely with the planned trajectories generated by the navigation system. Here are the key components and steps proposed for this development:

1. **Controller Design**:
2. **Reinforcement Learning Framework**
3. **Integration with Existing Systems**

Reference

Xiao, X., Dhamankar, G., Nair, A., et al. (2020). Adaptive Planner Parameter Learning from Reinforcement (APPLR): Enhancing Navigation Systems Through Reinforcement Learning. University of Texas at Austin.

Fujimoto, S., van Hoof, H., & Meger, D. (2018). Addressing Function Approximation Error in Actor-Critic Methods.

Perille, D., Truong, A., Xiao, X., & Stone, P. (2020). Benchmarking Metric Ground Navigation.

Silver, D., Lever, G., Heess, N., et al. (2014). Deterministic Policy Gradient Algorithms.

Lillicrap, T. P., Hunt, J. J., Pritzel, A., et al. (2016). Continuous control with deep reinforcement learning.

Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement Learning: A Survey. Journal of Artificial Intelligence Research.

Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning: An Introduction (2nd ed.). MIT Press.Brockman, G., Cheung, V., Pettersson, L., et al. (2016). OpenAI Gym.