IJARSCT

ISSN (Online) 2581-9429

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 1, September 2024

Learning Adaptive Control of a UUV Using a Bio-Inspired Experience Replay Mechanism

Dr. Pradeep V, Mr. Sreedeep P, Ms. Srusti Vaibhav, Ms. Sneha, Mr. Somesh K H
Department of CSE (IoT, Cyber Security including BlockChain)
Alva's Institute of Engineering and Technology, Mijar, Karnataka, India
pradeepv@aiet.org.in, sreedeeppothan@gmail.com, srustivaibhav09@gmail.com,
snehaacharya552@gmail.com, somesh.k.h2609@gmail.com

Abstract*: This paper provides an in-depth analysis of the present state of Deep Reinforcement Learning (DRL) applications in Unmanned Underwater Vehicles (UUVs). Addressing the persistent challenges related to data inefficiency and performance degradation in physical platforms, particularly when faced with unforeseen variations, the paper introduces the innovative Biologically-Inspired Experience Replay (BIER) method. This approach incorporates two distinct memory buffers to enhance learning efficiency. The paper assesses the generalization capabilities of BIER through training neural network controllers on diverse tasks, spanning from inverted pendulum stabilization to simulating half-cheetah running. Furthermore, BIER is integrated with the Soft Actor-Critic (SAC) method for UUV stabilization under unknown environmental dynamics. Evaluation in a ROS-based UUV simulator, incorporating increasingly complex scenarios, showcases BIER's superior performance over traditional Experience Replay (ER) methods, achieving optimal UUV control in half the time. This review contributes valuable insights into the challenges and advancements in applying DRL methods to UUVs, highlighting the BIER method's promising potential to improve adaptability and efficiency in UUV manoeuvring tasks, leading to more robust and agile underwater vehicle control systems for more robust and agile underwater vehicle control systems. .*

Keywords: Unmanned Underwater Vehicles

I. INTRODUCTION

This study addresses the conventional limitation of Unmanned Underwater Vehicle (UUV) autopilots, which predominantly rely on velocity and orientation sensors to compensate for disturbances induced by waves and currents. Current UUV autopilots effectively handle low-frequency components of sea-induced disturbances, but their performance can be enhanced by considering the nature of these disturbances in the autopilot design. The proposed approach leverages adaptive control within the framework of learning-based methods, where machine learning algorithms address the unknown or unmodeled aspects of a process, complementing traditional control methods that manage the known part. Disturbances in the UUV environment, such as marine currents, are treated as the unknown component, while the known component corresponds to the vehicle's maneuverability in the absence of disturbances. The paper presents the innovative Bio-Inspired Replay (BIER) method, which integrates principles from the biological replay mechanism into Deep Reinforcement Learning (DRL) algorithms. BIER extends the traditional Experience Replay (ER) strategy by introducing two new buffers: the sequential partial memory, utilizing incomplete state-action pair sequences to train the algorithm with an emphasis on recent policies, and the optimistic memory, highlighting positive reinforcement by increasing the probability of transitions associated with high-reward regions during ER training.

II. LEARNING-BASED ADAPTIVE CONTROL

In the realm of learning-based adaptive controllers for non-linear systems with uncertainties, model-free algorithms address the lack of a complete process model. These algorithms, including Deep Reinforcement Learning (DRL) methods, approximate unknown parts of the model, while classical model-based control efficiently handles known components. DRL, defined as a Markov Decision Process, involves an agent interacting with an environment to maximize

DOI: 10.48175/IJARSCT-19508

rewards through deep neural networks approximating states and actions. Deep Policy Gradient (DPG) algorithms, particularly the Soft Actor-Critic (SAC) method, have proven effective in adapting control parameters for robotic tasks. The learning process involves action selection, reward receipt, and value updates based on the learned policy. Experience Replay (ER) mechanisms, proposed in this work, aim to mitigate distribution shift issues, enhancing the performance of DPG methods sensitive to such challenges.

III. EXPERIENCE REPLAY (ER)

The general Experience Replay (ER) method involves storing an agent's experiences in a replay buffer, from which mini-batches are randomly sampled to train Artificial Neural Networks (ANNs) approximating the optimal policy. Key parameters impacting ER performance include buffer size, age of a transition, and replay ratio. To address issues with buffer size, the Combined Experience Replay (CER) method adds the latest transition to the mini-batch, but it maylead to performance drops. This paper introduces a novel ER mechanism aiming to decouple agent performance from process complexity, resolving CER related issues. Insights from recent analysis suggest that increasing replay capacity and buffer size with fixed replay ratio enhance learning by reducing overfitting. The proposed mechanism incorporates biological experience replay principles, considering temporally structured replay, reward modulation, and selective replay weighted by novelty. This novel ER mechanism integrates biological insights while addressing constraints related to the regression problem.

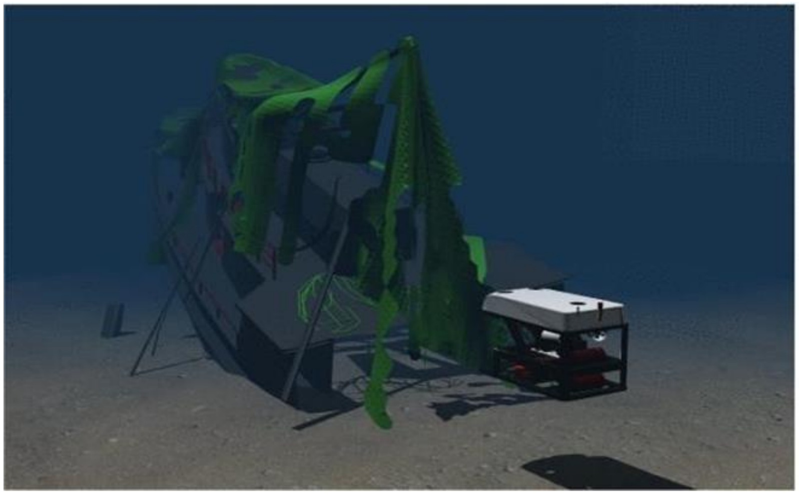
IV. UUV MANOEUVRING CONTROL

Discussion and Future Direction:

This study focuses on the control of Unmanned Underwater Vehicle (UUV) maneuvering tasks, specifically stabilizing the vehicle at a fixed velocity and orientation. The state vector, denoted as x, comprises position (xyz) and orientation angles (ϕ, θ, ψ). The fully actuated UUV is exposed to external disturbances, including first-order current- induced forces (zeromean oscillatory motions) and second-order wave-induced forces (nonzero varying components), assumed non-observable. The dynamics involve both known (f1) and unknown (f2) components.

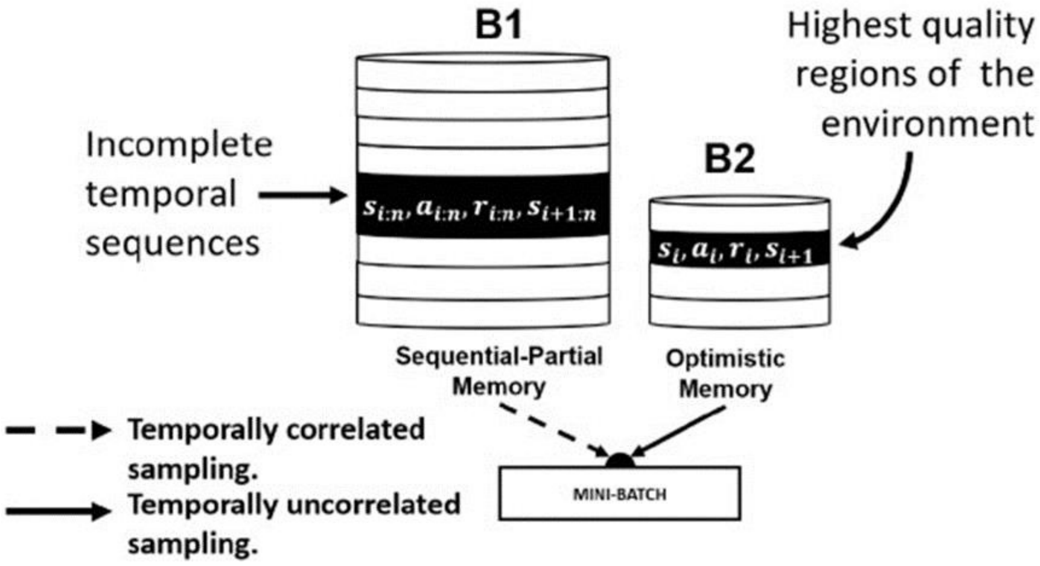
The control objective aims to steer the UUV to maintain the error signals (difference between present and desired states) within a specific threshold (χ) over a defined period, ensuring vehicle stabilization. This is expressed as the condition $\forall\ t' \in [t - \varsigma, t]$, $\nexists i \in Ru$ (control inputs), such that $|ei(t')| > \chi$. Here, t is the current time step, and ς is thetime period during which all errors ei remain below a small threshold χ.

The experimental platform used is the RexROV2, a cubic-shaped UUV detailed in [13], serving as the physicalmodel for the model-based part of the controller.



A Bio-Inspired Experience Replay (BIER):

In the presented Biologically-Inspired Experience Replay (BIER) method, two separate memory units are considered: the sequential-partial memory (B1), responsible for storing incomplete temporal sequences, and the optimistic memory (B2), which prioritizes high reward transitions based on the current policy. As depicted in Figure 2,BIER capitalizes on the robustness of on-policy sampling while retaining the effectiveness of data obtained through the off-policy formulation.


B1
Highest quality regions of the environment
Incomplete temporal sequences
$s_{i:n}, a_{i:n}, r_{i:n}, s_{i+1:n}$
B2
s_i, a_i, r_i, s_{i+1}
Sequential-Partial Memory
Optimistic Memory
Temporally correlated sampling.
Temporally uncorrelated sampling.
MINI-BATCH

The Biologically-Inspired Experience Replay (BIER) method introduces two memory units: the sequential- partial memory (B1) for storing temporally correlated transitions and the optimistic memory (B2) emphasizing high- reward transitions. B1's maximum size is set to 1,000,000 items, sampled with a 1:2 transition ratio for regularization.B2 stores outliers of the reward distribution, considering a transition as an outlier if its reward surpasses the expected future rewards. B2's maximum size is 10,000, focusing on current best transitions. A mini-batch is constructed from both memory units. The objective is to use Soft Actor-Critic (SAC) for UUV control, and results are referred to as PID+BIER. Contrasting with Combined Experience Replay (CER) referred to as PID+CER, and the baseline results, PID, involve off-the-shelf PID controllers in UUVsimulations.

V. CONCLUSION

This study addresses challenges in controlling autonomous underwater vehicles (UUVs) under extreme conditions by integrating classical control with advanced machine learning. The approach combines classical control for the known part of the UUV's process with the SAC algorithm, a deep reinforcement learning method, for learning the unknown part representing environmental disturbances. The novel Biologically Inspired Experience Replay (BIER) method enhances SAC, demonstrating faster adaptability and improved stability in simulated complex environments. The computational efficiency of DRL suggests potential real-world application on standard onboard processors. This adaptive control method, blending classical control robustness with learning-based adaptability, holds promise for advancing UUV applications. Future work will explore its generalization to diverse vehicles and environmental conditions.

REFERENCES

[1]. Mnih, V., Kavukcuoglu, K., Silver, D., et al. (2015). Human-level control through deep reinforcement learning. Nature, 518(7540), 529-533. DOI: 10.1038/nature14236.

[2]. Lillicrap, T.P., Hunt, J.J., Pritzel, A., et al. (2016). Continuous control with deep reinforcement learning. Proceedings of the International Conference on Learning Representations (ICLR).

[3]. Haarnoja, T., Zhou, A., Abbeel, P., & Levine, S. (2018). Soft Actor- Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. Proceedings of the 35th International Conference on Machine Learning (ICML).

[4]. Schaul, T., Quan, J., Antonoglou, I., & Silver, D. (2015). Prioritized Experience Replay. Proceedings of the International Conference on Learning Representations (ICLR).

[5]. Riedmiller, M. (2005). Neural fitted Q iteration – first experiences with a data efficient neural reinforcement learning method. European Conference on Machine Learning (ECML). Springer, Berlin, Heidelberg. Fujimoto, S., van Hoof, H., & Meger, D. (2018). Addressing Function Approximation Error in Actor-Critic Methods. Proceedings of the 35th International Conference on Machine Learning (ICML).

[6]. Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning: An Introduction. 2nd ed., MIT Press.

[7]. Levine, S., Pastor, P., Krizhevsky, A., Ibarz, J., & Quillen, D. (2018). Learning Hand-Eye Coordination for Robotic Grasping with Deep Learning and Large-Scale Data Collection. The International Journal of Robotics Research, 37(4-5), 421-436. DOI: 10.1177/0278364917710318.

[8]. Kober, J., Bagnell, J.A., & Peters, J. (2013). Reinforcement Learning in Robotics: A Survey. The International Journal of Robotics Research, 32(11), 1238-1274. DOI: 10.1177/0278364913495721.

[9]. Berger, M., Atanasov, N., LaValle, S. M., & Kumar, V. (2020). Autonomous quadrotor navigation using reinforcement learning. Robotics and Autonomous Systems, 133, 103641. DOI: 10.1016/j.robot.2020.103641.

[10]. Recht, B., & Recht, S. (2019). A Tour of Reinforcement Learning: The View from Continuous Control. Annual Review of Control, Robotics, and Autonomous Systems, 2(1), 253-279. DOI: 10.1146/annurev-control- 053018-023825.

[11]. Thrun, S., Burgard, W., & Fox, D. (2005). Prob'abilistic Robotics. MIT Press.

[12]. Fossen, T. I. (1994). Guidance and Control of Ocean Vehicles. Wiley. Kumar, R., Jiang, L., Bekris, K.E., & Abbeel, P. (2019). Learning navigation policies from demonstrations using dynamic trajectory templates. Proceedings of the Conference on Robot Learning (CoRL).

[13]. Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement Learning: A Survey. Journal of Artificial Intelligence Research, 4, 237- 285.

DOI: 10.48175/IJARSCT-19508