AoI-MDP: An AoI Optimized Markov Decision Process Dedicated in The Underwater Task (Student Abstract)

Yimian Ding[1*], Jingzehua Xu[1*], Yiyuan Yang[2], Guanwen Xie[1], Xinqi Wang[1], Shuai Zhang[3]
[1]MicroMasters Program in Statistics and Data Science, Massachusetts Institute of Technology, USA
[2]Department of Computer Science, University of Oxford, United Kingdom
[3]Department of Data Science, New Jersey Institute of Technology, USA
yimiandingthu@gmail.com, xjzh23@berkeley.edu, yiyuan.yang@cs.ox.ac.uk, gwxie360@outlook.com,
aaahayooha@163.com, sz457@njit.edu

Abstract

Ocean exploration places high demands on autonomous underwater vehicles, especially when there's observation delay. We propose age of information optimized Markov decision process (AoI-MDP) to enhance underwater tasks by modeling observation delay as signal delay and including it in the state space. AoI-MDP also introduces wait time in the action space and integrates AoI with reward functions, optimizing information freshness and decision-making using reinforcement learning. Simulations show AoI-MDP outperforms the standard MDP, demonstrating superior performance, feasibility, and generalization in underwater tasks. To accelerate relevant research, we have made the codes available as open-source at https://github.com/Xiboxtg/AoI-MDP.

Introduction

Utilizing autonomous underwater vehicles (AUVs) with reinforcement learning (RL) is a significant research focus for ocean exploration (Xu et al. 2024b; Li et al. 2020; Wu et al. 2022). However, tasks often fail due to observation delays caused by acoustic communication, leading to non-causality in control policies (Ding et al. 2024; Wei et al. 2023; Yi et al. 2020). Most studies use the standard Markov decision process (MDP) without accounting for observation delay, assuming instant state information reception(Xu et al. 2024a; Talak, Karaman, and Modiano 2020). This idealization is impractical due to signal propagation delays and channel congestion, which reduce information freshness and decision-making efficiency (Altman and Nain 1992; Yates et al. 2021).

Based on above analysis, we propose an age of information (AoI) optimized MDP (AoI-MDP) for underwater tasks to improve performance with observation delay. Our contributions include the following:

- To the best of our knowledge, we are the first to model underwater tasks as an MDP with observation delay and AoI, using RL for AUV training to optimize information updating and decision-making strategies jointly.
- We utilize statistical signal processing (SSP) for high-precision modeling of observation delay via AUV equipped sonar, yielding realistic results.
- Comprehensive experiments in the underwater data collection task show AoI-MDP's superior feasibility and performance in balancing multi-objective optimization.

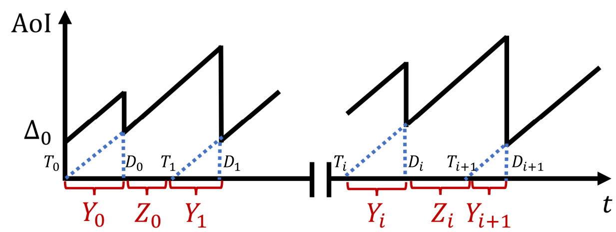


Figure 1: Illustration of the AoI model.

Methodology

AoI Optimized Markov Decision Process. As illustrated in Figure 1, consider the i-th underwater acoustic signal is transmitted from the AUV at time T_i, and the corresponding observed information is received at time D_i, AoI is defined using a sawtooth piecewise function

$$\Delta(t) = t - T_i, D_i \leq t < D_{i+1}, \forall i \in \mathbf{N}. \quad (1)$$

In AoI-MDP, AoI is not only part of reward functions to guide objective optimization via RL, but also crucial side information for decision-making. The AoI-MDP's state space, action space, and reward functions are detailed as follows:

State Space $\boldsymbol{\mathcal{S}}$: consists of the AUV's observed information $s_i^{'}$ and observation delay Y_i at time i, represented by $s_i = (s_i^{'}, Y_i) \in \boldsymbol{\mathcal{S}}^{'} \times \boldsymbol{Y}$. High-precision modeling of $s_i^{'}$ and Y_i is achieved via SSP, whose details are presented later.

Action Space $\boldsymbol{\mathcal{A}}$: composed of the tuple $a_i = (a_i^{'}, Z_i) \in \boldsymbol{\mathcal{A}}^{'} \times \boldsymbol{Z}$, where $a_i^{'}$ denotes the actions taken by the AUV and Z_i indicates the wait time between observing the environmental information and decision-making at time i.

Reward Function $\boldsymbol{\mathcal{R}}$: Apart from the original reward function $r_i^{'}$ in standard MDP, AoI-MDP introduces the time-averaged AoI as a new reward component. Thus, the updated reward function can be represented by tuple $r_i = (r_i^{'}, -\bar{\Delta})$. And the time-averaged AoI can be computed as follows: $\bar{\Delta} = \frac{\sum_{i=1}^{\mathcal{N}}((2Y_{i-1}+Y_i+Z_{i-1})\times(Y_i+Z_{i-1}))+S_0}{2\times(\sum_{i=1}^{\mathcal{N}} Z_{i-1}+\sum_{i=1}^{\mathcal{N}} Y_i+Y_0)}$, where $\mathcal{N}$ is the length of information signal, $S_0 = 0.5 \times (2\Delta_0 + Y_0) \times Y_0$.

*These authors contributed equally.

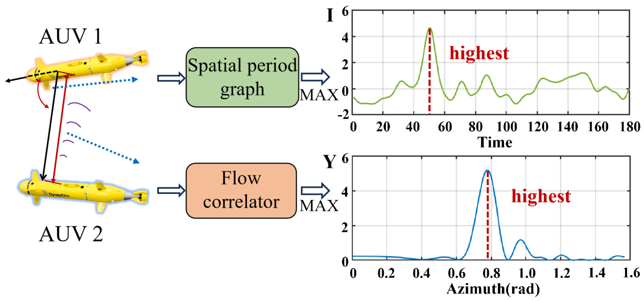


Figure 2: Diagram of the azimuth and time delay estimation.

Observation Delay and Information Modeling. Different from previous work, our study enhances the state space of AoI-MDP by incorporating observed information from AUV sensors and treating observation delay as underwater acoustic signal delay, as shown in Figure 2.

Specifically, our study assumes the AUV uses a sonar system to estimate distances to environmental objects. We further employ the flow correlator as an estimator to determine the time delay, which can be represented as

$$\mathcal{X}[n] = \mathcal{S}[n{-}Y_i] + \mathcal{W}[n], n = 0, 1, \ldots, N-1, \tag{2a}$$

$$J[Y_i] = \sum_{n=Y_i}^{Y_i+M-1} \mathcal{X}[n]\mathcal{S}[n-Y_i], 0 \le Y_i \le N-M, \tag{2b}$$

$$\hat{Y}_i = \operatorname{argmax}\left[J[Y_i]\right], \tag{2c}$$

where $\mathcal{S}[n]$ represents the known signal, while Y_i denotes the time delay to be estimated, and $\mathcal{W}[n]$ is Gaussian white noise with variance σ^2. where M is sampling length of $\mathcal{S}[n]$.

On the other hand, the AUV in our study utilizes a long linear array sensor to estimate the azimuth β between its orientation and environmental objects. The estimator in SSP is further leveraged to estimate the azimuth β. By maximizing the spatial period graph, the estimate of β ($0<\beta<\pi/2$) can be calculated as follows:

$$x\,[n] = Acos\left[2\pi\left(F_0\frac{d}{c}cos\beta\right)n+\phi\right]+\mathcal{W}[n], \quad n = 0, 1, \cdots, M-1, \tag{3a}$$

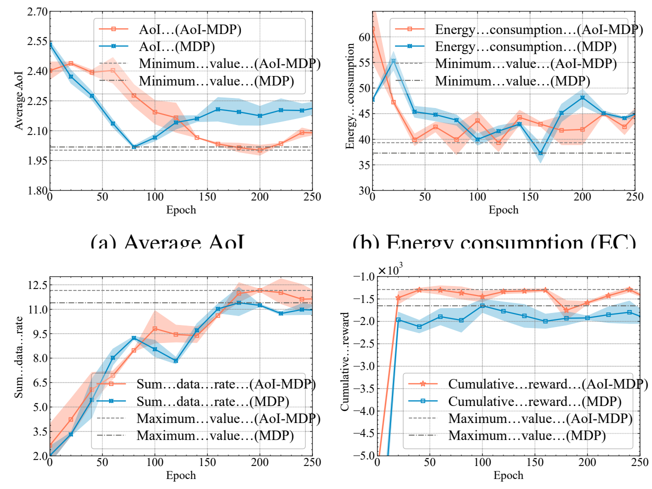
$$I_s(\beta) = \frac{1}{M}\left(\Big|\sum_{n=0}^{M-1} x\,[n] \exp\left[-j2\pi(F_0\frac{d}{c}cos\beta)n\right]\Big|\right)^2, \tag{3b}$$

$$\hat{\beta} = \operatorname{argmax}\left[I_s\left(\beta\right)\right], \tag{3c}$$

where F_0 denotes the frequency of transmitted signal, while d represents the interval between sensors. Besides, c indicates the speed of underwater acoustic signal propagation.

Simulation Experiments

Since open-source underwater tasks are scarce, we use a multi-AUV data collection task to evaluate the AoI-MDP's feasibility and effectiveness. For more details and parameters, refer to the previous work (Zhang et al. 2024).



(a) Average AoI (b) Energy consumption (EC)

(c) Sum data rate (SDR) (d) Cumulative reward

Figure 3: Comparison of experimental results of RL training based on AoI-MDP and standard MDP.

	AoI	SDR	EC
SSP	**1.97±0.26**	**11.99±0.73**	**33.83±2.59**
Poisson	3.42±0.18	5.95±2.42	34.27±7.98
Exponential	2.67±0.26	7.65±1.99	43.11±4.16
Geometric	2.38±0.28	12.34±0.79	58.15±9.49

Table 1: Comparison of different delay models.

We compared RL training results based on AoI-MDP and standard MDP under identical conditions, respectively. As shown in Figure 3, AoI-MDP achieved lower time-averaged AoI, reduced energy consumption, higher sum data rate, and greater cumulative rewards, indicating improved training effectiveness and performance.

Then we assessed AoI-MDP's generalization using common delay models (exponential, poisson, geometric) and compared the results with the SSP model, as shown in Table 1. AoI-MDP demonstrated superior performance across various distributions, highlighting its strong generalization capabilities. The SSP model achieved near-optimal results in AoI, data rate, and energy consumption optimization, proving effective in the underwater data collection task.

Finally, given the limited space, we have made the codes and supplementary materials available as open-source at our arXiv version: https://arxiv.org/abs/2409.02424.

Conclusion

In this work, we introduce the AoI-MDP into the underwater task. By integrating wait time into the action space and linking AoI with the reward function, AoI-MDP enhances both information freshness and decision-making processes through reinforcement learning. Simulations confirm the AoI-MDP's superior performance in underwater tasks.

Ethical Statement

This paper presents work whose goal is to explore using AoI-MDP to enhance the performance of marine robotics for the underwater tasks. There are many potential societal consequences of our work, none of which we feel must be specifically highlighted here.

Acknowledgments

Part of this work was done when Yimian Ding and Jingzehua Xu were studying in the MicroMasters Program in Statistics and Data Science at Massachusetts Institute of Technology (MIT). We are very grateful to Yiyuan Yang and Shuai Zhang at University of Oxford and New Jersey Institute of Technology (NJIT) for their strong support, and to Miao Liu and Songtao Lu at IBM research and MIT-IBM Watson AI Lab and for their valuable advice, respectively. Additionally, we thank all anonymous reviewers for their constructive comments.

References

Altman, E.; and Nain, P. 1992. Closed-loop control with delayed information. *SIGMETRICS Perform. Eval. Rev.*, 20(1): 193–204.

Ding, Y.; Wang, X.; Xu, J.; Xie, G.; Liu, W.; and Li, Y. 2024. Multi-Objective-Optimization Multi-AUV Assisted Data Collection Framework for IoUT Based on Offline Reinforcement Learning. *arXiv preprint arXiv:2410.11282*.

Li, Y.; Liu, L.; Yu, W.; Wang, Y.; and Guan, X. 2020. Noncooperative Mobile Target Tracking Using Multiple AUVs in Anchor-Free Environments. *IEEE Internet of Things Journal*, 7(10): 9819–9833.

Talak, R.; Karaman, S.; and Modiano, E. 2020. Optimizing Information Freshness in Wireless Networks Under General Interference Constraints. *IEEE/ACM Transactions on Networking*, 28(1): 15–28.

Wei, W.; Wang, J.; Du, J.; Fang, Z.; Ren, Y.; and Chen, C. L. P. 2023. Differential Game-Based Deep Reinforcement Learning in Underwater Target Hunting Task. *IEEE Transactions on Neural Networks and Learning Systems*, 1–13.

Wu, J.; Song, C.; Ma, J.; Wu, J.; and Han, G. 2022. Reinforcement Learning and Particle Swarm Optimization Supporting Real-Time Rescue Assignments for Multiple Autonomous Underwater Vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 23(7): 6807–6820.

Xu, J.; Ding, Y.; Zhang, Z.; Xie, G.; Wang, Z.; Zeng, Y.; and Li, G. 2024a. Multi-AUV Assisted Seamless Underwater Target Tracking Relying on Deep Learning and Reinforcement Learning. In *2024 International Joint Conference on Neural Networks (IJCNN)*, 1–9.

Xu, J.; Zhang, Z.; Wang, Z.; Wang, J.; and Rent, Y. 2024b. VoI and Energy-Aware AUV-Assisted Data Collection for Internet of Underwater Things. In *2024 IEEE Wireless Communications and Networking Conference (WCNC)*, 1–6.

Yates, R. D.; Sun, Y.; Brown, D. R.; Kaul, S. K.; Modiano, E.; and Ulukus, S. 2021. Age of Information: An Introduction and Survey. *IEEE Journal on Selected Areas in Communications*, 39(5): 1183–1210.

Yi, M.; Wang, X.; Liu, J.; Zhang, Y.; and Bai, B. 2020. Deep reinforcement learning for fresh data collection in UAV-assisted IoT networks. In *IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 716–721. IEEE.

Zhang, Z.; Xu, J.; Xie, G.; Wang, J.; Han, Z.; and Ren, Y. 2024. Environment and Energy-Aware AUV-Assisted Data Collection for the Internet of Underwater Things. *IEEE Internet of Things Journal*, 11(15): 26406–26418.