Adaptive Switching for Multimodal Underwater Acoustic Communications Based on Reinforcement Learning

Cheng Fan and Zhaohui Wang
Dept. of Electrical and Computer Engineering, Michigan Technological University
Houghton, Michigan, USA
fancheng@mtu.edu,zhaohuiw@mtu.edu

Abstract

The underwater acoustic (UWA) channel is a complex and stochastic process with large spatial and temporal dynamics. This work studies the adaptation of the communication strategy to the channel dynamics. Specifically, a set of communication strategies are considered, including frequency shift keying (FSK), single-carrier communication, and multicarrier communication. Based on the channel condition, a reinforcement learning (RL) algorithm, the Depth Determined Strategy Gradient (DDPG) method along with a Gumbel-softmax scheme is employed for intelligent and adaptive switching among those communication strategies. The adaptive switching is performed on a transmission block-by-block basis, with the goal of maximizing a long-term system performance. The reward function is defined based on the energy efficiency and the spectral efficiency of the communication strategies. Simulation results reveal that the proposed method outperforms a random selection method in time-varying channels.

Keywords: multimodal, underwater acoustics, DDPG, adaptive communication

ACM Reference Format:
Cheng Fan and Zhaohui Wang. 2021. Adaptive Switching for Multimodal Underwater Acoustic Communications Based on Reinforcement Learning. In *The 15th International Conference on Underwater Networks I& Systems (WUWNet'21), November 22–24, 2021, Shenzhen, Guangdong, China.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3491315.3491354

1 Introduction

The underwater acoustic (UWA) channel is a complex and stochastic process which exhibits large spatial and temporal dynamics. Effective underwater acoustic communication requires adaptation of the communication strategy to the constantly varying channel conditions. This work studies adaptive switching among different communication strategies based on the channel condition through reinforcement learning (RL).

The last two decades have witnessed considerable progress on various techniques for UWA communications, particularly the transceiver design for both single-carrier and multicarrier communications. Those techniques exhibit different properties. For instance, the single-carrier technique is more suitable for the UWA channel with a small delay spread, while the multicarrier technique can better handle the large channel delay spread but is constrained by the channel Doppler spread [1]. Therefore, intelligent switching among those techniques according to the channel condition is expected to yield better communication performance.

The RL is a learning paradigm where an agent takes a series of intelligent actions in an unknown environment, and through environmental exploration and the exploitation of learned knowledge to maximize a long-term reward [4]. The RL has been used in recent works for adaptive modulation and coding (AMC) in UWA communications [5]. Specifically, a model-based RL was proposed in [5] for adaptive transmission to maximize the energy efficiency.

2 Adaptive Switching for Multimodal UWA communications

2.1 Problem Formulation

This work considers the blocked UWA transmission. Within each block, the UWA channel is assumed quasi-stationary. The system can switch among different communication strategies on a block-by-block basis.

Without loss of generality, the discrete action space is formed by three communication strategies and two discrete transmission power levels, as listed in Table I. A multicarrier technique, the orthogonal frequency-division multiplexing (OFDM) technique is used.

Define N_{bits} as the number of information bits in each transmission block. Define T_{block} as the block time duration, and define B as the system bandwidth. Define P_{block} as the transmission power. Define p_{s} as the average bit success rate in each block transmission. The spectral efficiency R_{s} and the energy efficiency R_{e} can be formulated,

WUWNet'21, November 22–24, 2021, Shenzhen,Guangdong, China

ACM ISBN 978-1-4503-9562-5/21/11.
https://doi.org/10.1145/3491315.3491354

Table 1. The Discrete Action Space

Actions	Communication Strategy	Power level
1	2FSK	1
2	Single-carrier (QPSK)	1
3	OFDM (QPSK)	1
4	2FSK	2
5	Single-carrier (QPSK)	2
6	OFDM (QPSK)	2

respectively, as $R_s := (N_{\text{bits}} \times p_s)/(B \times T_{\text{block}})$, and $R_e := (P_{\text{block}} \times T_{\text{block}})/(N_{\text{bit}} \times p_s)$.

The reward function is defined as

$$R := \alpha_1 \times \tanh'(R_s) + \alpha_2 \times \tanh'(1/R_e), \quad (1)$$

where α_1 and α_2 are weighting coefficients, and the hyperbolic tangent function $\tanh' := \tanh(ax + b)$, introduces the non-linearity to emphasize the workable range of p_s in practical systems, with a and b being turning parameters.

2.2 The Proposed Method

Based on the channel condition, an policy-based RL algorithm, the Depth Deter-mined Strategy Gradient (DDPG) method [3] is employed for intelligent and adaptive switching among different communication strategies. Four neutral networks, including an actor network, a target actor network, a critic network and a target critic network, are used in DDPG. Consider the discrete action space in Table 1. Reparameterization is required to sample the original continuous action space of DDPG. However, direct sampling is not differentiable, which is undesirable for the gradient-based RL method. Here, a Gumbel-Softmax scheme [2] is introduced to replace the direct sampling and the argmax function in DDPG to make each step differentiable. For details, please refer to [2].

3 Simulation Results

The proposed method is evaluated in the Additive White Gaussian Noise (AWGN) channel. Two noise levels are introduced to simulate the channel dynamics. We consider 60 time blocks in total. The channel noise level in first 20 blocks is 9.29 dB, while in the last 40 blocks is 11.53 dB. The operational frequency band is [21, 27] kHz. The bit rate of the OFDM and of the single-carrier is 6 kbit/s, and the bit rate of 2FSK is 1.5 kbit/s. For the reward function R, related coefficients are set as $\alpha_1 = \alpha_2 = 1$, $a = 40$ and $b = -0.95$.

We compare the proposed method with a random selection method in which each action has an equal probability to be selected. 10 Monte Carlo simulations are used. Performance of those two methods along with the theoretical optimal performance are shown in Fig. 1. One can see a gap between the performance of the proposed method and the optimal performance, since neutral networks are used in DDPG and it is not guaranteed that the proposed method can converge to the optimal solution in every trail. However, the proposed method still tends to converge to the optimal solution in the first 20 blocks, whose performance is superior than the average performance of the random selection method. After the noise level change, the proposed method converges to a new policy closing to the optimal solution.

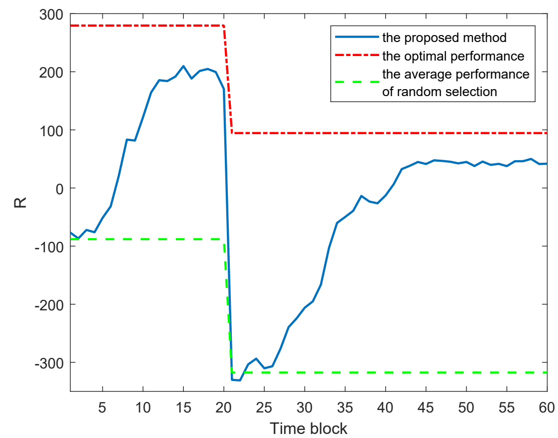

Figure 1. Immediate reward of the proposed method and the random selection in AWGN channels. In the first 20 blocks, the optimal action is #5; in the last 20 blocks, the optimal action is #4.

4 Conclusions

This work studied adaptive switching for multimodal UWA communications. A DDPG algorithm augmented by the Gumbel-Softmax scheme was developed to find the action for each time block that can maximize a long-term system performance. Simulation results in AWGN channels showed that the proposed method achieves better performance than the random selection method when the channel state changes. In the future work, we will evaluate the proposed method in practical time-varying UWA channels.

Acknowledgments

This work was supported by the NSF grant ECCS-1651135.

References

[1] A.R. Bahai, B.R. Saltzberg, and M. Ergen. 2004. *Multi-carrier digital communications: theory and applications of OFDM.* Springer Science I& Business Media, Boston.

[2] E. Jang, S. Gu, and B. Poole. 2017. Categorical reparametrization with Gumble-softmax. In *Proceedings of the ICLR Conference.*

[3] T.P. Lillicrap and et al. 2015. Continuous control with deep reinforcement learning. arXiv:1509.02971

[4] R.S. Sutton and A.G. Barto. 2018. *Reinforcement learning: An introduction.* MIT press, Massachusetts.

[5] C. Wang, Z. Wang, W. Sun, and D.R. Fuhrmann. 2017. Reinforcement learning-based adaptive transmission in time-varying underwater acoustic channels. *IEEE access* 6, 5 (Dec. 2017), 2541–2558. https://doi.org/10.1109/ACCESS.2017.2784239