MetaABR: Environment-Adaptive Video Streaming System with Meta-Reinforcement Learning

Wangyu Choi and Jongwon Yoon
{wangyu92,jongwon}@hanyang.ac.kr
Hanyang University, Korea

ABSTRACT

This work focuses on a video bitrate algorithm that quickly adapts to new and various environments with just a few update steps. This aspect is especially important for large-scale video streaming services used by a wide variety of users in different environments. Our proposed model is based on a neural network and employs meta-reinforcement learning to train it. After training, it can be easily customized for a variety of new environments with a few update steps, providing a user-specific streaming service.

CCS CONCEPTS

• **Networks → Transport protocols**; • **Computing methodologies → Machine learning**.

KEYWORDS

Adaptive Bitrate Algorithm, Meta-reinforcement Learning

ACM Reference Format:
Wangyu Choi and Jongwon Yoon. 2022. MetaABR: Environment-Adaptive Video Streaming System with Meta-Reinforcement Learning. In *CoNEXT Student Workshop 2022 (CoNEXT-SW '22), December 9, 2022, Roma, Italy.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3565477.3569155

1 INTRODUCTION

The COVID-19 pandemic has been reshaping our lifestyle, especially video streaming and conferencing services have become more prevalent. With the development of network infrastructure, bandwidth shortages are being alleviated, however, video streaming environments are becoming increasingly complex with a variety of high-quality content, heterogenous mobile devices, unpredictable mobility patterns, etc. In such a complex environment, it is difficult to provide a satisfactory service to all users with a generalized streaming solution. In order to guarantee high quality of experience (QoE) to users, many data-driven, machine learning-based bitrate adaptation (ABR) algorithms have been proposed recently [1–3].

Despite these efforts, state-of-the-art works still have several limitations in large-scale video streaming services. First, the ABR algorithm with a single strategy is difficult to provide high-quality services to all users. A large-scale video streaming system caters to users from various environments, and different strategies must be adopted accordingly for high QoE. Second, the state-of-the-art neural network-based models have to go through multiple update steps to revise its strategy. Since the video streaming environment in the real world is constantly changing over time, the ABR algorithm must be able to quickly adapt its strategy to the user's environment in real time. Third, fixed rule-based and traditional machine learning-based algorithms make it difficult to change strategies according to user preferences. Some users prefer high-quality video, and some users prefer seamless video, which means that it is difficult for a single strategy to satisfy all users.

In this work, we present MetaABR, an ABR algorithm that quickly adapts to a new environment by employing meta-reinforcement learning. The training process of MetaABR consists of two parts: configuring various environments and finding transferable parameters. After deployment, MetaABR adapts to the new environment with just a few updates from the transferable model. This allows real-time adjustment of model policies in a highly variable environment, as well as the ability to make decisions that reflect user preferences.

2 RELATED WORK

Various machine learning-based ABR solutions have been proposed for different environments. Pensieve [1], which generates a policy directly with reinforcement learning, has made improvements in terms of QoE. Oboe+MPC [2] modifies the policy by tuning the parameters of the ABR algorithm in real time. Fugu [3] trains a model to predict the transmission time of future segments and uses MPC to determine the bitrate. Although they have made significant performance gains, they only provide an average policy for large-scale video streaming services.

3 SYSTEM DESIGN

The goal of MetaABR is to provide an individually tailored ABR solution by adapting algorithm to various/new environments. Toward this, as depicted in Figure 1, we establish three steps: (i) configuring a wide variety of environments, (ii) finding transferable parameters through intensive training, and (iii) adapting to new environments in the real world.

Step 1: configuring the environment. Before generating our model, we configure diverse and vast simulation-based environments. First, we enumerate the properties that create/affect the video streaming environment. Then we configure the environments by sampling the values of their characteristics. We carefully select properties as they constitute an environment that requires different strategies according to their diversity. We use the following properties to configure the environments: bandwidth, video, client's configuration (buffer size) and QoE coefficients. Finally, each property is combined to form a different video streaming environment.

CoNEXT-SW '22, December 9, 2022, Roma, Italy

ACM ISBN 978-1-4503-9937-1/22/12.
https://doi.org/10.1145/3565477.3569155

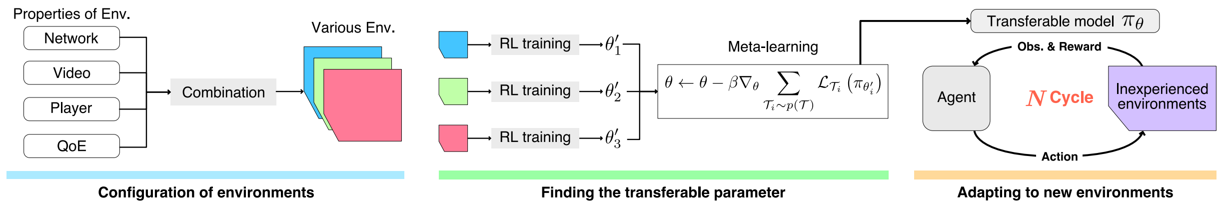


Figure 1: Three steps of MetaABR.

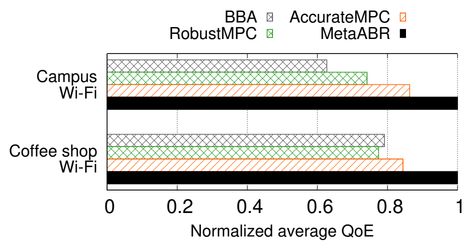


Figure 2: MetaABR outperforms state-of-the-art algorithms.

Step 2: finding transferable parameter. Now, we find a parameter θ that allows fast adaptation to the new environments by training in different environments. We adopt a Model-Agnostic Meta-Learning (MAML) algorithm [4] that combines the models obtained after training in different environments. Specifically, our training process consists of an internal adaptation and an external adaptation phases. The internal adaptation phase aims to compute the updated parameter θ_i' for the task $\mathcal{T}_i$ sampled from the distribution for the tasks $p(\mathcal{T}_i)$ (i.e., environments). For the internal update, we adopt Proximal Policy Optimization (PPO) [5], a state-of-the-art policy gradient algorithm. In the external adaptation phase, the parameter θ is updated according to the training trajectories of the PPO agent collected in the internal adaptation phase. Both phases alternate to update the parameter θ that produces the highest reward after the PPO agent is trained on the N update steps.

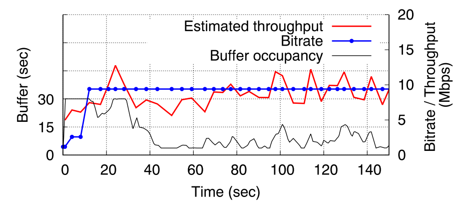
Step 3: adapting to new real-world environments. After obtaining the transferable parameter θ, the model is easily adapted to the new environment with only N update steps after deployment. The training agent updates θ to θ^* to adapt to unexperienced network, video, player configuration, and QoE preferences in just N-step updates (shown in Figure 1).

4 EVALUATION

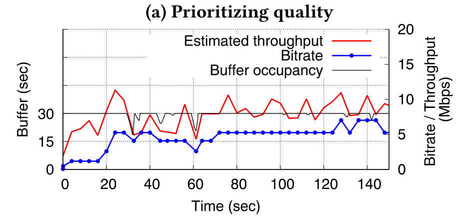
We implement MetaABR using dash.js and create the following two scenarios for evaluation.

Coping with new environment and video. Figure 2 shows the average QoE per segment using public Wi-Fi (coffee shop and campus), and Bug Buck Bunny video (i.e., new environments). The coffee shop public Wi-Fi is variable with bandwidth sharing by other users' traffic, and the campus Wi-Fi has sharp bandwidth drops due to handover of access points because we collected while walking. We use BBA [6], RobustMPC [7], and AccurateMPC as state-of-the-art algorithms for comparison. Due to implementation issues of Fugu [3], we use AccurateMPC, replacing Fugu's TTP (transmission time predictor) with accurate future bandwidth. We can see that MetaABR provides better streaming services on Wi-Fi with very variable or sharp drops.

Adjusting QoE preference. We scrutinize MetaABR's decisions for two different QoE preferences: quality preferred and seamless playback preferred (Figure 3). When prioritizing video quality, we can see that MetaABR maintains the bitrate despite its low buffer occupancy. On the other hand, MetaABR tries to keep the buffer occupancy high enough by adjusting the bitrate according to the estimated bandwidth.



(a) Prioritizing quality



(b) Prioritizing rebuffering penalty

Figure 3: MetaABR provides diverse adaptations based on user QoE preferences.

5 CONCLUSION AND FUTURE WORK

In this work, we propose MetaABR, an ABR algorithm that quickly adapts to various/new environments using meta-reinforcement learning. We demonstrate that simulation-based offline trained MetaABR in various environments outperforms existing algorithms in new environments never experienced before. Therefore, we provide the feasibility of an environment-adaptive solution in a large-scale video streaming service. In a future work, we extend the existing system to deploy it to a streaming service in the real world and analyze its intrinsic characteristics carefully. Specifically, we discuss why our approach is effective, in what situations it struggles, what are the optimal update steps, etc.

ACKNOWLEDGMENT

This work was supported by Basic Science Research Program through the National Research Foundation of South Korea (NRF) funded by the Ministry of Education NRF-2022R1A2C1008743 and the MSIT(Ministry of Science and ICT), Korea, under the Grand Information & Communication Technology Research Center support program (IITP-2022-2020-0-101741) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation).

REFERENCES

[1] Hongzi Mao, Ravi Netravali, and Mohammad Alizadeh. 2017. Neural Adaptive Video Streaming with Pensieve. In *Proceedings of the Conference of the ACM Special Interest Group on Data Communication (SIGCOMM '17).* Association for Computing Machinery, New York, NY, USA, 197—210. https://doi.org/10.1145/3098822.3098843

[2] Zahaib Akhtar, Yun Seong Nam, Ramesh Govindan, Sanjay Rao, Jessica Chen, Ethan Katz-Bassett, Bruno Ribeiro, Jibin Zhan, and Hui Zhang. 2018. Oboe: Auto-Tuning Video ABR Algorithms to Network Conditions. In *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication (SIGCOMM '18).* Association for Computing Machinery, New York, NY, USA, 44—58. https://doi.org/10.1145/3230543.3230558

[3] Francis Y. Yan, Hudson Ayers, Chenzhi Zhu, Sadjad Fouladi, James Hong, Keyi Zhang, Philip Levis, and Keith Winstein. 2020. Learning in situ: a randomized experiment in video streaming. In *Proceedings of the 17th Usenix Conference on Networked Systems Design and Implementation (NSDI'20).* USENIX Association, USA, 495—512.

[4] Chelsea Finn, Pieter Abbeel, and Sergey Levine. 2017. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70 (ICML'17).* JMLR.org, 1126—1135.

[5] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, Oleg Klimov. 2017. Proximal Policy Optimization Algorithms. arXiv:1707.06347. Retrieved from https://arxiv.org/abs/1707.06347

[6] Te-Yuan Huang, Ramesh Johari, Nick McKeown, Matthew Trunnell, and Mark Watson. 2014. A Buffer-Based Approach to Rate Adaptation: Evidence from a Large Video Streaming Service. In *Proceedings of the 2014 ACM Conference on Special Interest Group on Data Communication (SIGCOMM '14).* Association for Computing Machinery, New York, NY, USA, 187—198. https://doi.org/10.1145/2619239.2626296

[7] Xiaoqi Yin, Abhishek Jindal, Vyas Sekar, and Bruno Sinopoli. 2015. A Control-Theoretic Approach for Dynamic Adaptive Video Streaming over HTTP. In *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication (SIGCOMM '15).* Association for Computing Machinery, New York, NY, USA, 325—338. https://doi.org/10.1145/2785956.2787486