Sequential Information Design: Markov Persuasion Process and Its Efficient Reinforcement Learning

JIBANG WU, University of Virginia, USA
ZIXUAN ZHANG, University of Science and Technology of China, China
ZHE FENG, Google, USA
ZHAORAN WANG, Northwestern University, USA
ZHUORAN YANG, Yale University, USA
MICHAEL I. JORDAN, University of California, Berkeley, USA
HAIFENG XU, University of Chicago, USA

In today's economy, it becomes important for Internet platforms to consider the sequential information design problem to align its long term interest with incentives of the gig service providers (e.g., drivers, hosts). This paper proposes a novel model of sequential information design, namely the Markov persuasion processes (MPPs), in which a sender, with informational advantage, seeks to persuade a stream of myopic receivers to take actions that maximize the sender's cumulative utilities in a finite horizon Markovian environment with varying prior and utility functions. Planning in MPPs thus faces the unique challenge in finding a signaling policy that is simultaneously persuasive to the myopic receivers and inducing the optimal long-term cumulative utilities of the sender. Nevertheless, in the population level where the model is known, it turns out that we can efficiently determine the optimal (resp. ϵ-optimal) policy with finite (resp. infinite) states and outcomes, through a modified formulation of the Bellman equation that additionally takes persuasiveness into consideration.

Our main technical contribution is to study the MPP under the online reinforcement learning (RL) setting, where the goal is to learn the optimal signaling policy by interacting with with the underlying MPP, without the knowledge of the sender's utility functions, prior distributions, and the Markov transition kernels. For such a problem, we design a provably efficient no-regret learning algorithm, the Optimism-Pessimism Principle for Persuasion Process (OP4), which features a novel combination of both optimism and pessimism principles. In particular, we obtain optimistic estimates of the value functions to encourage exploration under the unknown environment, and additionally robustify the signaling policy with respect to the uncertainty of prior estimation to prevent receiver's detrimental equilibrium behavior. Our algorithm enjoys sample efficiency by achieving a sublinear $\sqrt{T}$-regret upper bound. Furthermore, both our algorithm and theory can be applied to MPPs with large space of outcomes and states via function approximation, and we showcase such a success under the linear setting.

The full paper is available at https://arxiv.org/abs/2202.10678

CCS Concepts: • **Theory of computation** → **Algorithmic game theory**; **Convergence and learning in games**; **Multi-agent learning**; **Sequential decision making**; *Regret bounds*; *Markov decision processes*.

Additional Key Words and Phrases: Strategic Learning, Information Design, Bayesian Persuasion, Sequential Decision Making, Reinforcement Learning

EC '22, July 11–15, 2022, Boulder, CO, USA.

ACM ISBN 978-1-4503-9150-4/22/07.
https://doi.org/10.1145/3490486.3538313

ACM Reference Format:
Jibang Wu, Zixuan Zhang, Zhe Feng, Zhaoran Wang, Zhuoran Yang, Michael I. Jordan, and Haifeng Xu. 2022. Sequential Information Design: Markov Persuasion Process and Its Efficient Reinforcement Learning. In *Proceedings of the 23rd ACM Conference on Economics and Computation (EC '22), July 11–15, 2022, Boulder, CO, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3490486.3538313

ACKNOWLEDGMENTS

Zhaoran Wang acknowledges National Science Foundation (Awards 2048075, 2008827, 2015568, 1934931), Amazon, J.P. Morgan, and Two Sigma for their supports. Haifeng Xu is supported by an NSF grant CCF-2132506.