香港中文大學（深圳）
The Chinese University of Hong Kong, Shenzhen

Modelling Competitive Behaviors in Autonomous Driving Under Generative World Model

Guanren Qiao, Guorui Quan, Rongxiao Qu, Guiliang Liu

Background

- Previous simulation systems typically predict vehicles' trajectories by imitating realistic driver behaviors from an offline dataset . While this approach can precisely replicate common occurrences, it often falls short in capturing Out-of-Distribution (OoD) and long-tail events.
- modern self-driving agents are capable of managing routine traffic scenarios; however, they often fail to handle safety-critical events in the long tail end of data distributions.

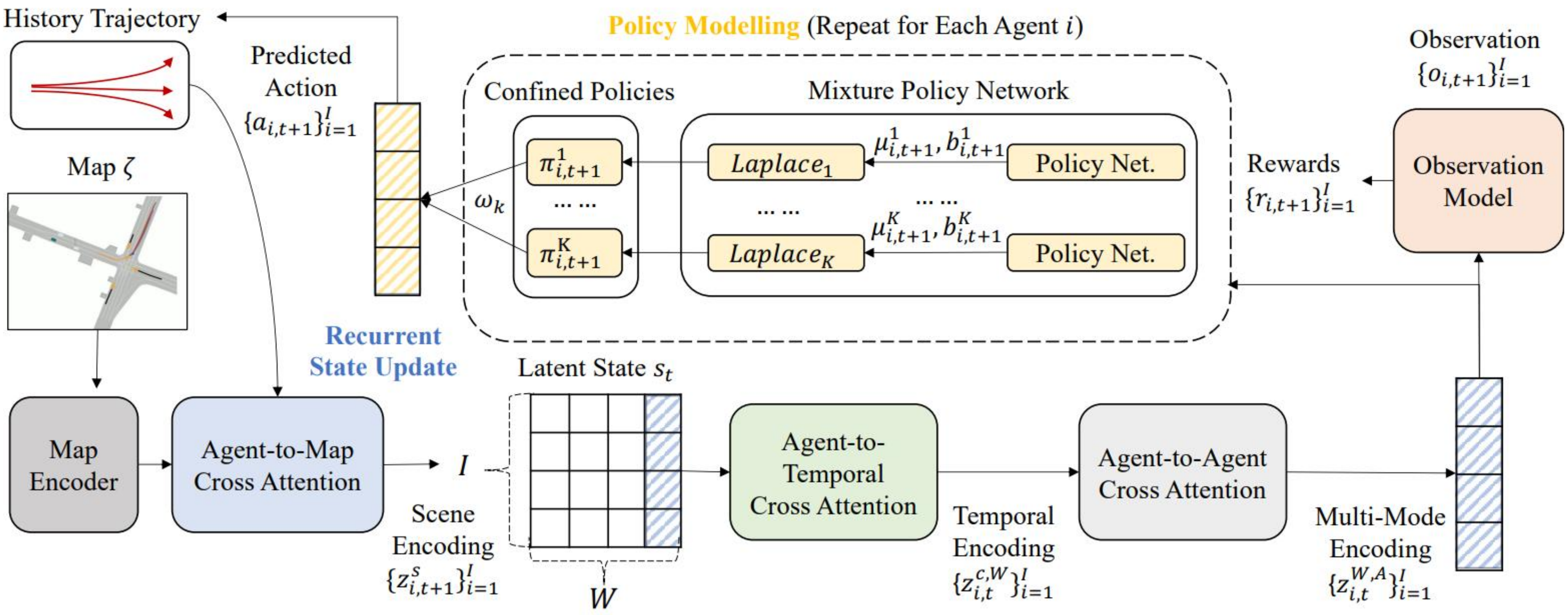
Contribution

- We develop a game-theoretic approach to generate critical events by viewing traffic simulation as a general-sum Markov game involving multiple agents.
- We integrate both 1) the optimistic Bellman Update and 2) the magnetic mirror decent into our learning objective to learn Coarsed Correlated Equilibirum (CCE) efficiently.
- We demonstrate the capability of the generative world model in enabling optimal control with exploitablitily experiment.
- We visualize the simulated traffic in several complex and competitive scenarios to understand CCE.

Problem Formulation

- Our problem setting is Offline Multi-agent Reinforcement Learning in General Sum Markov Games, which is based on the Dec-POMDP
- General Correlated Policy
- ϵ-approximate CCE

Methodology



Methodology

Identifying CCE from Multi-player General Sum Markov Game

1. Optimistic value function

$$\bar{V}_{i,t}^{\pi_i,\boldsymbol{\pi}_{-i}}(s) = \mathbb{E}_{\mu_0,p_T,\pi_i,\boldsymbol{\pi}_{-i}}\left[\sum_{\iota=t}^{T}\gamma^{\iota}\left[r_i^{\text{opti}}(\boldsymbol{o}_\iota,\boldsymbol{a}_\iota)\right]|\boldsymbol{o}_0=\mathcal{O}(s)\right]$$

2. Magnetic Mirror Descent

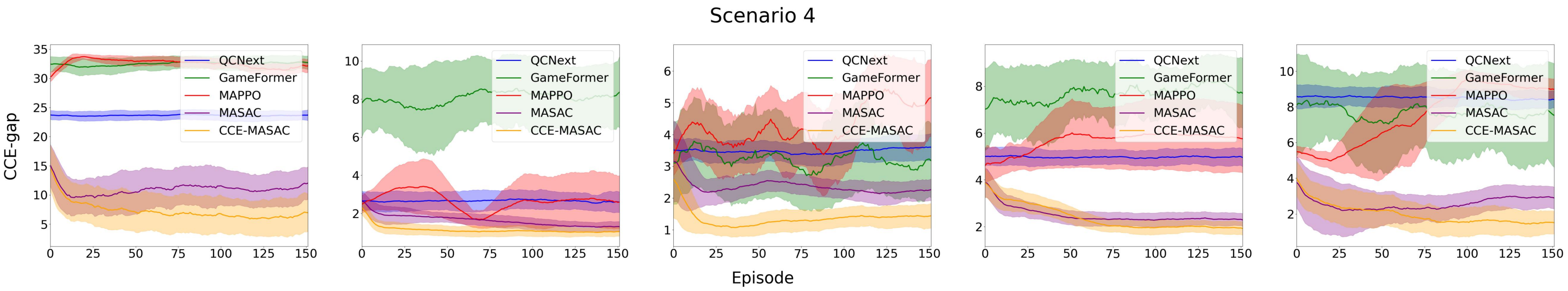
Our objective:

$$\max_{\pi_i\in\Pi_i^c}\mathbb{E}\Big[\sum_{t=0}^{T}\gamma^t\Big(r_i^{\text{opti}}(\boldsymbol{o}_t,\boldsymbol{a}_t)-\eta_1\mathcal{D}_{kl}[\pi_{i,t}(\cdot)\|\pi_{i,t}^{o}(\cdot)]-\frac{1}{\eta_2}\mathcal{D}_{kl}[\pi_{i,t}(\cdot)\|\pi_{i,t}^{j-1}(\cdot)]\Big)\Big]$$

$$\max_{\pi_i\in\Pi_i^c}\mathbb{E}\Big[\sum_{t=0}^{T}\gamma^t\Big(r_i^{\text{opti}}(\boldsymbol{o}_t,\boldsymbol{a}_t)+\eta\mathcal{H}[\pi_{i,t}(\cdot)]+\mathbb{E}_{\pi_{i,t}}[\log((\pi_{i,t}^{o})^{\eta_1}(\pi_{i,t}^{j-1})^{\frac{1}{\eta_2}}]\Big)\Big]$$

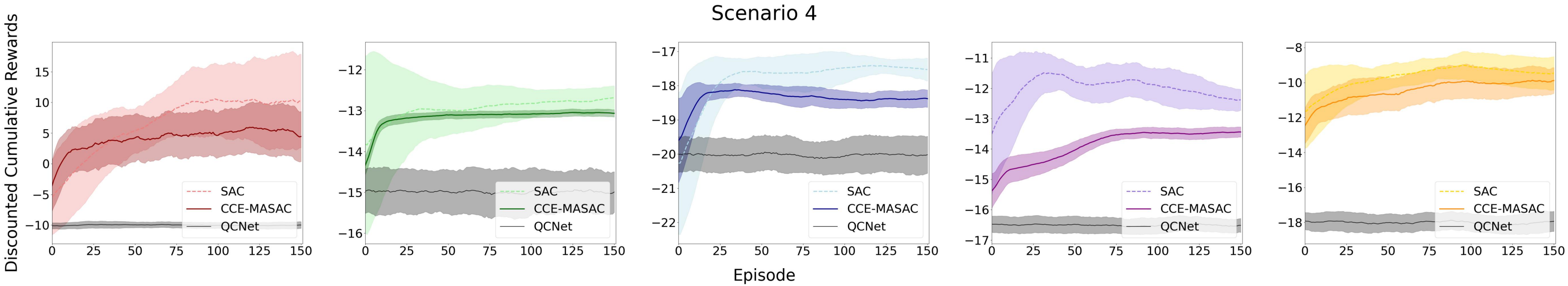
Experiments

Efficiency of Algorithms in Closing the CCE-gap



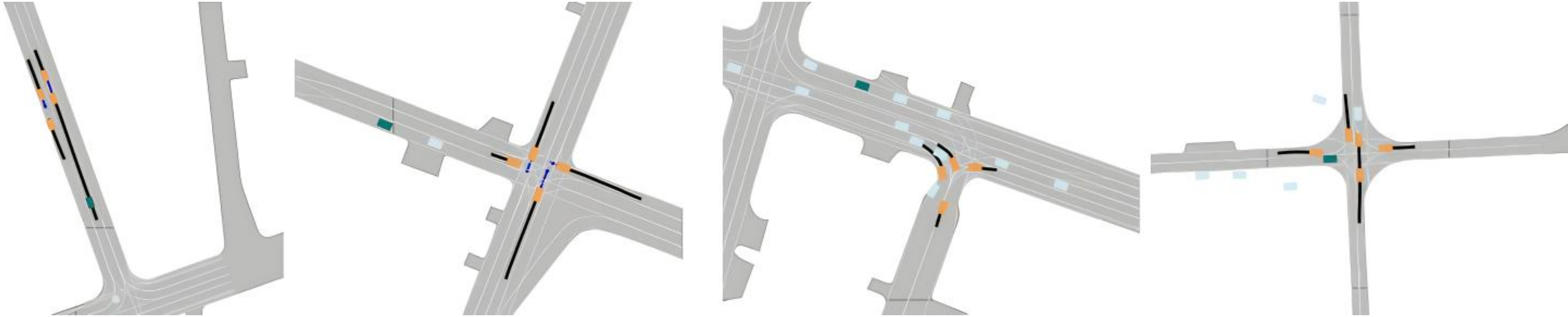
Experiments

Efficacy of the World Model in Facilitating Policy Update



Experiments

Case Study: Visualization of the Learned CCEs



Limitations

Omitting the Constraint: CCE-MASAC does not account for the constraints of agents, failing to model their behavior in avoiding constraint violations within competitive environments, which frequently occurs in real-world traffic situations.

Exclusion of Cooperation: While we focus on modeling the competitive behaviors of vehicles, realistic traffic scenarios might involve both competition and cooperation behaviors among agents during some specific events.

THE END

THANKS