Inverse Constrained Reinforcement Learning

Usman Anwar, Shehryar Malik, Alireza Aghasi, Ali Ahmed

INFORMATION TECHNOLOGY UNIVERSITY
CACTUS
ICML International Conference On Machine Learning

Introduction & Motivation

- Constraints are physically possible actions that should be avoided, for example, driving the car above the speed limit or breaking a traffic signals are examples of constraint violations.
- In any realistic environment (e.g. home), there are **too many constraints to enumerate**.
- Focus of this work: automated learning of constraints.

Contributions

- Provide a **model free** constraint learning method for **high dimensional, continuous setting**.
- Empirically show that **learned constraints transfer** to agents with different dynamics and morphologies.

Training Objective

Neural Network based soft parametrization of indicator set over constrained trajectories.

$$\frac{1}{N}\sum_{i=1}^{N}\nabla_\theta \log \zeta_\theta(\tau^{(i)}) - \frac{1}{M}\sum_{j=1}^{M}\nabla_\theta \log \zeta_\theta(\hat{\tau}^{(j)}) - \delta \sum_{\tau\sim\{\mathcal{D},\mathcal{S}\}} |1-\zeta_\theta(\tau)|$$

Samples From Expert — *Samples From Training Agent*

Objective can be loosely interpreted as trying to match average soft cost for both expert and RL agent.

Regularizer: Puts Penalty On Over Constraining.

Training Tricks

Importance Sampling

Sampling from training RL agent involves solving forward RL problem. That is expensive! Solution: use importance sampling!

$$\omega(s_t,a_t) = \frac{\zeta_\theta(s_t,a_t)}{\zeta_{\bar{\theta}}(s_t,a_t)}$$

Current Neural Network

Sampling Neural Network

KL Based Early Stopping

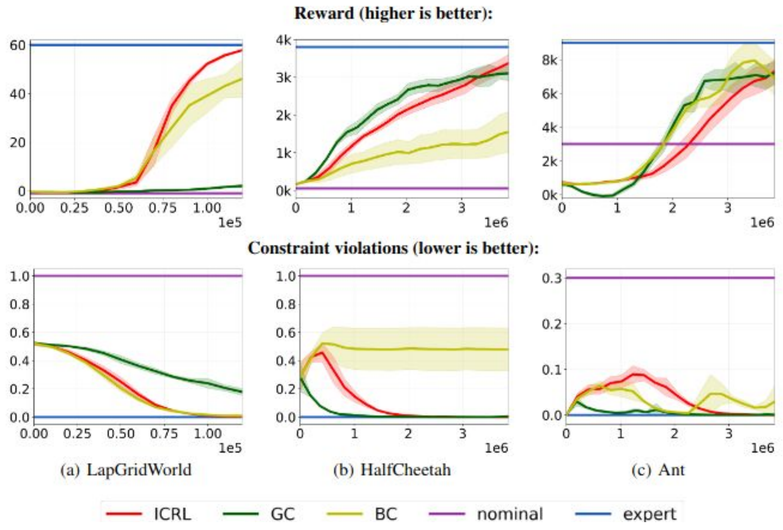
Use KL Based Early stopping to control for variance introduced by IS.

$$D_{\mathrm{KL}}(\pi_{\bar{\theta}}||\pi_\theta) \leq 2\log\bar{\omega}$$

$$D_{KL}(\pi_\theta||\pi_{\bar{\theta}}) \leq \frac{\mathbb{E}_{\tau\sim\pi_{\bar{\theta}}}\left[(\omega(\tau)-\bar{\omega})\log\omega(\tau)\right]}{\bar{\omega}}$$

Comparison With Baselines

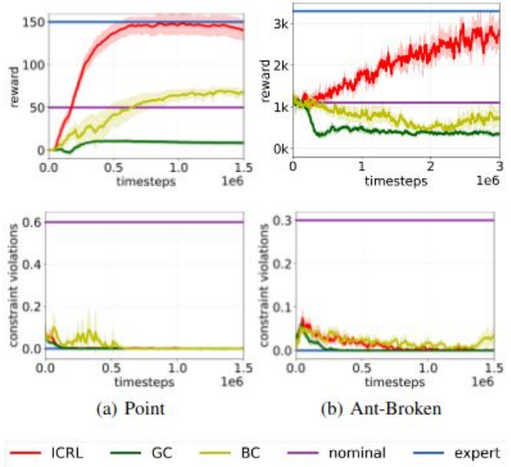
We benchmark the algorithm against a *binary classifier baseline* and a *GAIL inspired baseline*.



Transfer Experiments

Question: Can constraint net obtained from demonstration of one agent transfer to other agents?

Answer: Yes. Even when new agents differ in morphology or have different dynamics.



Ablation Study

Question: What components of the algorithm are critical to its performance?

Answer: All of them!