Developing Ethical Agents via Context-Sensitive Modular Inverse Reinforcement Learning

Boshi Wang, Yue Guo, Dana Hughes, Katia Sycara

ShanghaiTech University, Carnegie Mellon University

Introduction

Background:

Autonomous robots will soon enter our everyday lives and interact with people as co-workers and assistants, so it is important for these agents to be able to **reason about social norms and moral values under varying contexts to gain human acceptability**.

Main Problem of Reinforcement Learning Framework [1]:

hand-crafting appropriate rewards for each type of ethical/non-ethical behaviors is very challenging.

Our Approach & Highlights:

- **Inverse reinforcement learning:** recover reward function from expert demonstrations; avoid hand-tuning rewards.
- **Context-Sensitive Modular decision process:** capable of handling environment with **large number of changing contexts.**

Problem Formulation & Motivation

A set of contextual expert trajectories are generated from a **Context-Sensitive Decision Process** where each trajectory is

$$\tau = \{(s_1, c_1, a_1), ..., (s_T, c_T, a_T)\}$$

Goal: recover the reward function under which the optimal policy *π*(a|s,c)* induces behaviors that are similar to the expert's behavior.

Problems with current approaches

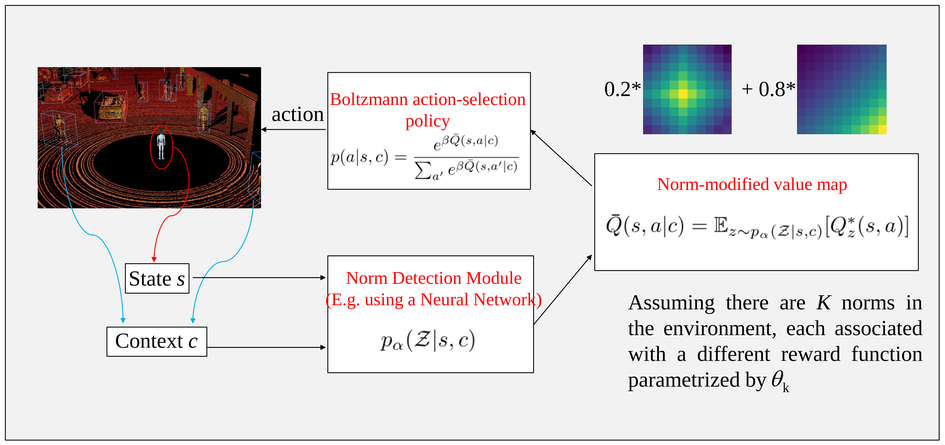
- No existing work tackles the computational challenge of ethics learning under large number of changing contexts.

Motivation behind our approach

- Although there are a lot of contexts, it is the ethical norms that govern the behaviors. The number of norms is usually a lot smaller than the number of contexts.
- In real situations, to a larger extent, we act on what we perceive (e.g. help someone who's in trouble), instead of predicting what future contexts would be all the time.

Methodology

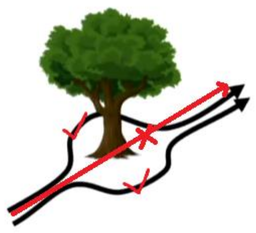
Context-Sensitive Modular Decision Process



Inverse Learning via Maximum Likelihood Estimate

Optimizing Norm-detection Module

$$\begin{aligned}\nabla_\alpha \log p(a_t|s_t,c_t) &= \beta \sum_{z_k} (Q^*_{z_k}(s_t,a_t) \\ &\quad - \sum_{a'_t} p(a'_t|s_t,c_t) Q^*_{z_k}(s_t,a'_t)) \nabla_\alpha p_\alpha(z_k|s_t,c_t) \\ &= \beta \sum_{z_k} A^{\pi_{\theta_k}}(s_t,a_t|c_t) \nabla_\alpha p_\alpha(z_k|s_t,c_t) \\ &= \beta \nabla_\alpha \mathbb{E}_{z_k\sim p_\alpha(z|s_t,c_t)}[A^{\pi_{\theta_k}}(s_t,a_t|c_t)]\end{aligned}$$



Maximizing expected conditional advantage of expert demonstrations

Optimizing reward parameters

For each θ_k:

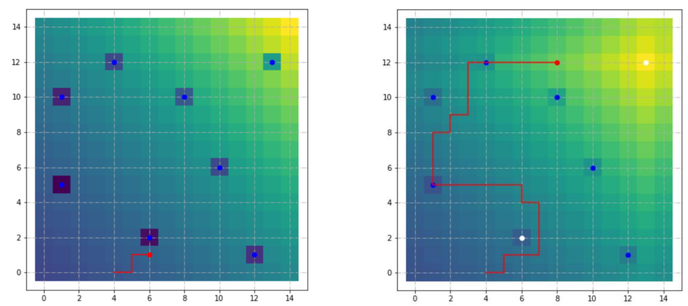
$$\begin{aligned}\nabla_{\theta_k} \log p(a_t|s_t,c_t) &= \beta p(z_k|s_t,c_t)(\nabla_{\theta_k} Q^*_{z_k}(s_t,a_t) \\ &\quad - \sum_{a'_t} p(a'_t|s_t,c_t) \nabla_{\theta_k} Q^*_{z_k}(s_t,a'_t))\end{aligned}$$

$\nabla_{\theta_k} Q^*_{z_k}(s,a)$ can be solved via fixed-point estimate:

$$\begin{aligned}\nabla_{\theta_k} Q^*_{z_k}(s,a) &= \nabla_{\theta_k} R_{\theta_k}(s) \\ &\quad + \gamma \mathbb{E}_{s'\sim p(s'|s,a)} \sum_{a'} \pi^*_{\theta_k}(s',a') \nabla_{\theta_k} Q^*_{z_k}(s',a')\end{aligned}$$

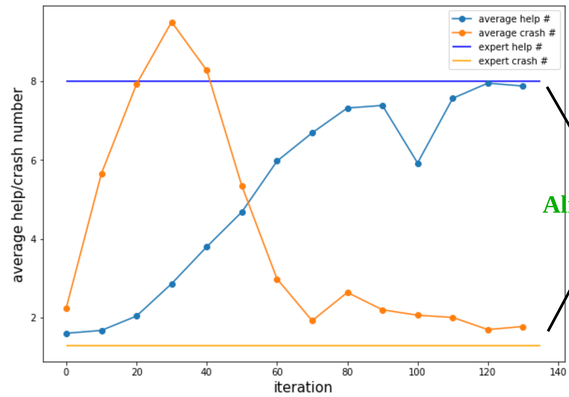
Experiments & Results

Scenario: Dynamic "Grab a Milk" (extension of [2])



We extend the scenario in [2] to a more challenging and realistic setting: now each baby's status is dynamic instead of fixed throughout, meaning that at each time step the baby may begin crying. Therefore the agent has to deal with **changing contexts in a large context space** (exponential in the number of babies in the environment).

Simulation Results



Aligned Value!

Conclusion & Future Work

We propose Context-Sensitive Modular IRL, a method for developing ethical agents which learns from expert demonstrations a set of reward functions that can induce ethical begaviors. Our model is capable of handling problems with very large context space and the complexity is linear in the number of norms in the environment.

In the future, we intend to implement the method in large scale real applications, extend the existing work to multi-agent scenarios and analyze the sample complexity of our method.

References

[1] D. Abel, J. MacGlashan, and M. L. Littman, "Reinforcement learning as a framework for ethical decision making," in Workshops at the Thirtieth AAAI Conference on Artificial Intelligence, 2016.
[2] Y.-H. Wu and S.-D. Lin, "A low-cost ethics shaping approach for designing reinforcement learning agents," in Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

Acknowledgment

Boshi Wang thanks Katia Sycara, Dana Hughes, Yue Guo for guidance, and ShanghaiTech University for funding the work. He also gives special thanks to Dr. John Dolan, Ms. Rachel Burcin and the RISS team for support.