

# Massively Multiagent Hierarchical Inverse Reinforcement Learning in Open-world Environments

William H. Guss\*  
wguss@cs.cmu.edu

Ruslan Salakhutdinov\*  
rsalakhu@cs.cmu.edu

January 17, 2018

## Abstract

**TODO:** Will, can you put together a quick 1.5-2 page proposal on inverse RL, Frank from US Army is saying that there is a potential to fund this kind of research: working with simulation environments, working on agents that can plan, communicate, solve slam, build strategies, do reward shaping via inverse RL, etc. They basically want to emphasize: 1, Metrics and Experiments. 2. Describing state-of-the-art in this area, and why what we are proposing is elite and new. Looks like US Army likes very ambitious projects. I will this together into one coherent white paper and we can see whether they would be willing to fund us. If not, we can take the same proposal to someone else as well.

## 1 Introduction

**TODO: Introduction to inverse reinforcement learning; what has it done?**

Over recent years inverse reinforcement learning (IRL) has revolutionized the state of the art in apprenticeship learning, cooperative and adversarial modeling, and the modeling of intent in human and animal behaviour**TODO: cite**. At its core, IRL solves the problem of learning expert policies indirectly: in direct juxtaposition to behavioral cloning IRL learns the reward function of an expert agent and then produces a policy which maximizes that reward. In this regime, the resultant policy is often far more interpretable, robust, and sample efficient than that of behavioural cloning.

Mechanically, inverse reinforcement learning is optimal not only for cloning expert policies in applications where thousands if not millions of demonstrations are not possible, but also as a substitute for traditional reinforcement learning when exploration is extremely expensive. For example, in robotic manipulation tasks, where typical  $\epsilon$ -greedy exploration policies would result in potential damage to the robot, apprenticeship learning via IRL is a powerful alternative. Furthermore, in forecasting tasks such as epidemiological modeling, traffic prediction, and **TODO: cite**

Formally, let  $R$

Due to the power of IRL, practitioners have been able to predict traffic patterns, I don't really know what I'm saying with this. So I think the best strategy is to write as much as possible and

---

\*Carnegie Mellon University, Machine Learning Department.

then cut. How about that. Bullshit whatever but spin this into a narrative that you'd tell Core to the success of IRL is not only its ability to recapitulate expert policies from a relatively low number of samples but also

**TODO: Provide a slightly more formal perspective to motivate why these methods could be interesting?**

**TODO: Introduce the problem statement, what is the problem of massively multi-agent hierarchical reinforcement learning in open-world environments and why would this be pertinent.**

**TODO: Introduce the implications of solving the MMHIRL problem in open-world environments, potentially reference the taxi problem.**

## 2 Our Approach

**TODO: To solve the problem of MMHIRL we will introduce methods which take advantage of the local connectedness of parameterized policy and reward space.**

**TODO:**

### 2.1 Experiments

### 2.2 Metrics

## References