Gaining Insights From Expert Demonstrations Using Inverse Reinforcement Learning

Max Pettersson[1,2]
Supervisors: Maria Riveiro[1,3] and Florian Westphal[1,3]

***Abstract*— This extended abstract presents a project that aims to study expert knowledge incorporation using Inverse Reinforcement learning (IRL). The purpose of this integration is for the IRL agent to find optimal solutions for problems where there is human expertise not embedded in the data and extract insights from the IRL solution that can improve the expert's knowledge.**

I. INTRODUCTION

For Reinforcement Learning (RL) problems, it is common to represent the environment as a Markov Decision Process (MDP), with a state space $\mathbf{S}$, action space $\mathbf{A}$, a transition function $\mathbf{P}(s'|s,a)$, a reward function $\mathbf{R}$, and a discount factor γ. The MDP is commonly represented as a tuple $M := (S, A, R, \gamma, P)$. More specifically, when the decision maker is in state $s \in S$, they choose an action $a \in A$ based on the current state. The MDP then probabilistically determines the next state $s' \in S$ and the reward $r \in R$ based on the current state s, action taken a, and the transition probabilities given by $\mathbf{P}$. Typically, an RL agent will search through states and perform actions that maximize the reward. A collection of actions to perform for each state is typically referred to as a *policy* π, and the policy that yields the highest expected cumulative reward is considered the *optimal policy* π^* [1].

Normally, an RL agent will try to find π^* for a reward function that is explicitly defined by its creator, and whether or not the agent will learn desired behaviors is highly dependent on how the reward function is defined. One base assumption for RL is that the reward function is static and known. This is not always the case. There are problem domains, for example, computational models for humans and animals, where the reward function is unknown and thus not able to be defined manually. In addition, a reward function for a particular environment may not be the same for the agent in a different environment [2]. This need to manually specify reward functions limits the applicability of RL in real-life problems. In these problems however, you often have demonstrated desired behaviors you could learn from, even if the reward function is unknown or too complex to design [3].

To address this, inverse Reinforcement Learning (IRL) was introduced. IRL is a collection methods that aims to identify a reward function given an observed behavior. Instead of using a manually defined reward function to search through the model of the MDP to find π^*, in IRL, a policy is demonstrated to the agent, and the agent searches for an optimal reward function R^* to explain the demonstration. Essentially the inverse of regular RL. IRL has some advantages; instead of searching through a very large or infinite policy space, the policy space is restricted to at least a good enough policy. Another advantage is that reward functions have been shown to be more generalizable than policies for transferring to new agents and environments [3]. The problem that IRL tries to solve is, however, an ill-posed problem. One of the base assumptions about the demonstrated policy is that it is optimal; still, there may be multiple different reward functions that may describe a demonstration [2], meaning it can be hard to guarantee that a reward function is optimal.

The demonstrations for an IRL agent can be conceptualized as providing the agent with prior knowledge of the problem. Von Rueden et al. [4] present a taxonomy that categorizes different knowledge types and how they have been integrated into machine learning models. The review describes three main sources of knowledge: *Scientific Knowledge*, which largely relates to mathematical representations of the world, *World Knowledge*, which encapsulates formally described knowledge like linguistics and semantics, and lastly *Expert Knowledge* which is described as a less formal application of world knowledge which just a small subset of the population knows. Most expert knowledge has been incorporated through probabilistic relations and human feedback. Meaning that an expert may be able to describe the results of their intuition, i.e., how one event might be more likely than another event (which may come from some type of intuition), and this could then inform how a learning model should weight its predictions. One unique aspect of expert knowledge is intuition that experts have gained from their experience within the domain. Human intuition is a concept that does not have a concise scientific definition, and thus is hard to quantify [5], but it may not need to be directly measured in order to be incorporated into a machine learning model. Instead, the expert could demonstrate how they would solve a problem, circumventing the need to formalize an expert's knowledge and intuition.

Explainable Artificial Intelligence (XAI) is a field of study focused on developing machine learning and artificial intelligence systems that are transparent and interpretable to humans. The goal of XAI is to make AI systems more accountable, understandable, and trustworthy by providing explanations for their decisions and actions. The need for

[1]Department of Computing, Jönköping University, Gjuterigatan 5, Jönköping, Sweden.
[2]max.pettersson@ju.se
[3]maria.riveiro@ju.se, florian.westphal@ju.se

XAI arises because modern AI systems, particularly, deep learning models, often work as a "black box," meaning that their decision-making processes are opaque to humans. As AI is increasingly integrated into various applications, including those with significant societal impact, it becomes essential to understand how AI systems arrive at their predictions. IRL has been used in conjunction with XAI to produce insights and strategies for human consumption. For example, for navigational instruction generation [6], allowing robots to communicate their objectives [7], and modeling optimal eye-gaze behavior for electric wheelchair drivers [8].

II. Problem Statement

The concept of prior knowledge has been leveraged for machine learning in the form of IRL by demonstrating to an agent a policy to follow and then letting the agent try to find a generalized reward function. This central concept of demonstrations carries an important but flawed assumption; that the demonstration shown to the agent is optimal. But if there is expert knowledge within the problem domain, and this expert knowledge can be formalized into a good demonstration for an IRL agent, it may prove useful to reinforce the assumption that the demonstration is optimal or at least make sure the model converges to a useful policy.

Using expert knowledge to provide demonstrations for an IRL model is an interesting potential solution to somewhat guarantee optimality of a demonstrated policy. This can lead to another interesting idea of using these machine learning models trained on expert knowledge to help train human novices, and generate insights to change training. If an IRL model can be somewhat guaranteed to have found an optimal solution for the learning task with expert knowledge, the insights that this model provides could be used to train humans for the same task. One way of extracting insights from models is using techniques from XAI to make the IRL model explain itself and its predictions.

Using machine learning to automate human training may be beneficial in areas where there is a shortage of human instructors, and in areas where the training may have to adapt based on a changing problem domain. An automated system for training may also provide some consistency based on empirical facts compared to a human expert that relies on their own experiences. A machine learning model, in the form of an RL agent, may also be able to find strategies and insights previously not discovered by humans that could improve the training of new experts. This is especially true for problem domains where most of the training doctrine is built upon previous experiences rather than empirical analysis.

III. Research Questions

From the problem description, the following research questions are formalized. The research questions are sequential and will be studied in the order presented.

1) How can expert knowledge be formalized into a demonstration for an IRL agent to provide a useful solution?
2) How can the solution from the IRL agent provide insights for training humans?

IV. Methods

In this project, we will leverage experts from two companies to build IRL agents for specific machine learning tasks related to human training and navigation generation. The companies will present industrial cases with data and experts, and these cases will be used as a basis for building machine learning tasks to study the research questions.

A literature review will be conducted to map out the techniques that have been used in related works. The first step will be for the IRL agent to find optimal solutions to the machine learning tasks and evaluate them with the experts. If the IRL agents have found optimal solutions with the help of expert knowledge, XAI techniques will be applied to find what insights the IRL agent has found. To evaluate the effectiveness of these insights, pseudo-experiments and surveys will be conducted using human participants. By analyzing the results of the experiments, we can gain a better understanding of how the insights discovered by the IRL agents can be used to improve training outcomes. The insights from the IRL agents may then provide some empirical basis for training improvements in a mostly human experience-driven training doctrine for these industrial cases.

Overall, this project has the potential to find new ways of solving problems related to human training and navigation generation, and provide valuable insights into how experts and novices in these domains can be trained more effectively. In addition, it will enhance the understanding of integrating prior knowledge into machine learning and broaden the problem areas where IRL has been applied.

References

[1] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Prentice Hall, 3 ed., 2010.

[2] A. Y. Ng and S. J. Russell, "Algorithms for inverse reinforcement learning," in *Proceedings of the Seventeenth International Conference on Machine Learning*, ICML '00, (San Francisco, CA, USA), p. 663–670, Morgan Kaufmann Publishers Inc., 2000.

[3] S. Arora and P. Doshi, "A survey of inverse reinforcement learning: Challenges, methods and progress," *Artificial Intelligence*, vol. 297, 8 2021.

[4] L. von Rueden, S. Mayer, K. Beckh, B. Georgiev, S. Giesselbach, R. Heese, B. Kirsch, J. Pfrommer, A. Pick, R. Ramamurthy, M. Walczak, J. Garcke, C. Bauckhage, and J. Schuecker, "Informed machine learning – a taxonomy and survey of integrating prior knowledge into learning systems," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 1, pp. 614–633, 2023.

[5] G. Lufityanto, C. Donkin, and J. Pearson, "Measuring intuition: nonconscious emotional information boosts decision accuracy and confidence," *Psychological science*, vol. 27, no. 5, pp. 622–634, 2016.

[6] A. F. Daniele, M. Bansal, and M. R. Walter, "Navigational instruction generation as inverse reinforcement learning with neural machine translation," vol. Part F127194, pp. 109–118, IEEE Computer Society, 3 2017.

[7] S. H. Huang, D. Held, P. Abbeel, and A. D. Dragan, "Enabling robots to communicate their objectives," *Autonomous Robots*, vol. 43, pp. 309–326, 2 2019.

[8] Y. Maekawa, N. Akai, T. Hirayama, L. Y. Morales, D. Deguchi, Y. Kawanishi, I. Ide, and H. Murase, "Modeling eye-gaze behavior of electric wheelchair drivers via inverse reinforcement learning," Institute of Electrical and Electronics Engineers Inc., 9 2020.