Symplex: Learning Social Norm Hierarchies by Combining Autonomous Exploration and Expert Imitation

Extended Abstract

Oliver Deane
University of Bristol
Bristol, United Kingdom
oliver.deane@bristol.ac.uk

Oliver Ray
University of Bristol
Bristol, United Kingdom
csxor@bristol.ac.uk

ABSTRACT

In this paper, we introduce SYMPLEX (Symbolic Policy Learning from Experts/Exploration), an interactive framework that learns complex hierarchies of behavioral norms as interpretable logical constraints through a combination of autonomous exploration and expert imitation. The approach ensures that learned constraints are interpretable for human oversight, generalizable for transfer to similar environments, and defeasible - enabling adaptation to novel behaviors and facilitating the learning of exceptions in dynamic domains. We demonstrate the utility of our approach in a traffic simulation environment using a neuro-symbolic implementation of SYMPLEX that interleaves a Deep Q-Learning (DQL) component for policy optimization through goal-directed domain exploration, with interactive Inductive Logic Programming (ILP) for example-based symbolic constraint generation. At each iteration, inferred constraints are imposed on the DQL via penalty terms appended to the reward function, allowing the system to form exceptions to previously-learned constraints. We illustrate SYMPLEX's ability to identify concise human-readable constraints in complex environments, and evidence the efficacy of learning norms as defeasible constraints. Additionally, we exemplify the benefits of using an interactive rule induction system in expediting convergence to accurate norms.

KEYWORDS

Multi-Agent Systems, Inverse Constrained Reinforcement Learning, Inductive Logic Programming, Interactive Machine Learning

ACM Reference Format:
Oliver Deane and Oliver Ray. 2025. Symplex: Learning Social Norm Hierarchies by Combining Autonomous Exploration and Expert Imitation: Extended Abstract. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Detroit, Michigan, USA, May 19 – 23, 2025*, IFAAMAS, 3 pages.

1 INTRODUCTION

Understanding and adhering to social norms is a fundamental requirement for social artificial agents operating in human-agent collaborative systems [2]. However, such norms are inherently nuanced and often implicit, making it impractical to fully encode them into an agent ab initio [11].

Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Y. Vorobeychik, S. Das, A. Nowé (eds.), May 19 – 23, 2025, Detroit, Michigan, USA. © 2025 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

An effective solution is Inverse Constrained Reinforcement Learning (ICRL), whereby goal-directed, exploration-driven learning is guided using constraints derived from expert examples [4, 5, 7, 9, 12]. ICRL methods alternate between updating an imitation policy and learning a constraint function from a set of expert demonstrations until convergence on a policy capable of reproducing expert behaviour. [5, 12]. A key limitation of current methods is their lack of interpretability, making it difficult to verify what an agent has learned. Ensuring safety requires validating which behaviors are extracted from demonstrations, but interpreting environmental constraints from numerous state-action pairs is challenging. Furthermore, learned constraints often struggle to transfer across environments since they are tied to specific state-action spaces.

Recent research highlights significant potential benefits in deriving social norms as human-readable symbolic constraints, generalized over individual states, using a logic-based machine learning method known as Inductive Logic Programming (ILP) [1]. Defining constraints as high-level, human-readable concepts enhances both interpretability and transferability. For example, instead of prohibiting specific off-road state-action pairs within a traffic environment, an agent can learn a general rule that driving off-road is undesirable [1]. This conceptual representation applies across diverse traffic environments more effectively than rigid state-based constraints.

However, at present, existing methods operate on a limited scale and are restricted to learning hard constraints that cannot be adjusted if updates to the set of expert trajectories elicit new, conflicting constraints. Further, these methods are built for fully-automatic deployment, and do not exploit interactive rule-induction methods that would further accelerate learning and reduce constraint violations while providing evidential usability benefits [3, 8, 10].

To address these issues, we propose a novel method for Symbolic Policy Learning from Examples/Exploration (SYMPLEX). As in ICRL, SYMPLEX learns by alternating between policy optimization and constraint generation. The policy optimizer is instantiated as a deep Q learner and constraints enforced at each iteration via a penalty term appended to the reward function. This enables operation in complex environments and the learning of defeasible social norms that can be selectively overridden when conflicting constraints arise. When the symbolic constraint generator is instantiated with an interactive ILP system, this enables interactive generation of interpretable constraints that can be iteratively refined with human input [10].

2 IMPLEMENTATION

SYMPLEX takes as input a set of expert trajectories $\boldsymbol{T}$ and a nominal (unconstrained) Markov Decision Process, and outputs a symbolic

representation of learned norms in the form of a logical hypothesis $\boldsymbol{H}$. All state-action pairs which are covered by $\boldsymbol{H}$ are considered as invalid and comprise the set of constraints $\boldsymbol{C}$.

It utilizes two sub-modules to iteratively build $\boldsymbol{H}$, and subsequently, $\boldsymbol{C}$. One sub-module consists of a Deep Q Learner (DQL) used to identify seed constraints - state action pairs deemed optimal according to the DQL's current learned policy, but are not contained within the set of expert trajectories. This is added to the set of overall constraints $\boldsymbol{C}$. ACUITY, an ILP system forming the partner induction module, generalizes this updated constraint set by inducing a hypothesis intended to represent the identified social norms [10]. All state-action pairs that satisfy this learned hypothesis are added to the overall set of constraints $\boldsymbol{C}$. $\boldsymbol{C}$ is then used to update the agent's nominal policy by applying graded penalties to the DQL's reward function, with newly learned constraints receiving larger associated penalties. A new seed constraint can then be inferred with this updated policy and the process repeats as $\boldsymbol{C}$ is iteratively augmented. Throughout this, end users can optionally exploit ACUITY's interactive mechanisms such as hypothesis shaping and example titration to avoid redundant clauses or to identify potential exceptions as new hypotheses are induced.

3 CASE STUDY: TRAFFIC SIMULATION

3.1 Environment

We validate the SYMPLEX method with a case study using the SUMO Traffic environment [6]. Within this environment, the behaviour of learning agents is constrained by accepted norms and rules of the road, namely: a) stay on the road, and b) only move into a junction on a green light.

3.2 Defeasible Constraint Inference

SYMPLEX successfully learns a logical program $\boldsymbol{H}$ that ensures it reaches defined goals while adhering to the environmental norms. A selected subset of the learned program is presented below, where A denotes an arbitrary state-action pair.

$$\begin{aligned} \text{invalid}(A) \ &:\text{-}\ \text{at}(A,B,C), \text{go}(A,D), \\ &\quad \text{move}(B,C,D,E,F), \neg\text{onRoad}(E,F). \\ \text{invalid}(A) \ &:\text{-}\ \text{at}(A,B,C), \text{go}(A,D), \\ &\quad \text{atJunction}(B,C,D), \text{traffic\_light}(A,red). \end{aligned} \tag{1}$$

As SYMPLEX learns this first constraint ("*stay on the road*"), it is enforced over the policy learner via coefficients within the reward function. As a result, learned constraints can be corrected and overridden if new contrastive constraints are introduced by the induction step. To demonstrate, we introduce an augmented set of expert trajectories $T+$ which represents a larger $10x10$ grid. Crucially, the northern section of this enlarged environment contains an obstacle on the road; here, we envision this as a pothole that the agent should avoid (see Figure 1). To navigate this, a learning agent would have to violate the learned *onRoad* constraint and leave the road, bypass the obstacle, and return to continue its journey to the goal.

Owing to SYMPLEX's graded penalization factors, the system can acquire a constraint that prevents the agent from entering the Pothole state without necessitating a complete re-initialization. (see Clause 2). The action of leaving the road remains permissible, all-be-it with a penalty, so the agent learns to violate the past constraint to continue on to its goal.

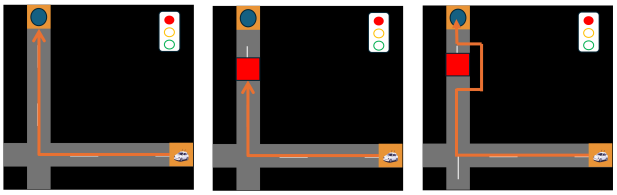

Figure 1: Pothole Scenario: the arrow depicts the agent's (car) trajectory to the goal. Left: an agent following the initial 'Stay on road' constraint. Middle: Non-defeasible hard constraints prohibit navigation around an obstacle (red square). Right: SYMPLEX's defeasible constraints allow a policy to override the previous constraint and proceed to the goal.

$$\begin{aligned} \text{invalid}(A) \ &:\text{-}\ \text{at}(A,B,C), \\ &\quad \text{beforePothole}(B,C,D), \text{not}(\text{go}(A,D)), \end{aligned} \tag{2}$$

3.3 Interactive Rule Induction

While SYMPLEX can be effective as a fully automatic system, there is evident benefits of exploiting ACUITY's interactive mechanisms to avoid redundant clauses or to identify potential exceptions. For example, a human-in-the-loop may intervene at the point at which the initial "stay on the road" clause was induced, and encourage ACUITY to search for an alternative rule which considers the "pothole" exception. This could be done through the titration of a potentially informative example taken from a wider dataset of unlabeled state-action pairs [10]. Alternatively, a user may rebut ACUITY's proposed clause and point it towards relevant sections of the hypothesis space that would elucidate a useful exception (e.g., clauses containing the $beforePothole/3$ predicate). In this way, ACUITY's interactive mechanisms can allow for an effective user-agent collaboration that accelerates the identification of reliable constraints.

4 CONCLUSION

In sum, we present a method for learning social norms in the form of high-level symbolic constraints by combining autonomous exploration with expert imitation. To accommodate exceptions to previously learned constraints, the system learns defeasible constraints through interactions with the agent's reward function, which can be shaped and refined by a human in the loop. As a result, the method generates norm exception hierarchies that are both interpretable and verifiable, with the necessary scalability to handle non-trivial examples.

REFERENCES

[1] Mattijs Baert, Sam Leroux, and Pieter Simoens. 2023. Inverse reinforcement learning through logic constraint inference. *Machine Learning* 112, 7 (2023), 2593–2618.
[2] Natalia Criado, Estefania Argente, and V Botti. 2011. Open issues for normative multi-agent systems. *AI communications* 24, 3 (2011), 233–264.

[3] Oliver Deane and Oliver Ray. 2023. Interactive model refinement in relational domains with inductive logic programming. In *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces*. 127–129.
[4] Ashish Gaurav, Kasra Rezaee, Guiliang Liu, and Pascal Poupart. 2022. Learning soft constraints from constrained expert demonstrations. *arXiv preprint arXiv:2206.01311* (2022).
[5] Guiliang Liu, Sheng Xu, Shicheng Liu, Ashish Gaurav, Sriram Ganapathi Subramanian, and Pascal Poupart. 2024. A Comprehensive Survey on Inverse Constrained Reinforcement Learning: Definitions, Progress and Challenges. *arXiv preprint arXiv:2409.07569* (2024).
[6] Pablo Alvarez Lopez, Michael Behrisch, Laura Bieker-Walz, Jakob Erdmann, Yun-Pang Flötteröd, Robert Hilbrich, Leonhard Lücken, Johannes Rummel, Peter Wagner, and Evamarie Wießner. 2018. Microscopic traffic simulation using sumo. In *2018 21st international conference on intelligent transportation systems (ITSC)*. IEEE, 2575–2582.
[7] Shehryar Malik, Usman Anwar, Alireza Aghasi, and Ali Ahmed. 2021. Inverse constrained reinforcement learning. In *International conference on machine learning*. PMLR, 7390–7399.
[8] Steve Moyle, Andrew Martin, and Nicholas Allott. 2024. XAI Human-Machine collaboration applied to network security. *Frontiers in Computer Science* 6 (2024), 1321238.
[9] Daehyung Park, Michael Noseworthy, Rohan Paul, Subhro Roy, and Nicholas Roy. 2020. Inferring task goals and constraints using bayesian nonparametric inverse reinforcement learning. In *Conference on robot learning*. PMLR, 1005–1014.
[10] Oliver Ray and Steve Moyle. 2021. Towards expert-guided elucidation of cyber attacks through interactive inductive logic programming. In *2021 13th International Conference on Knowledge and Systems Engineering (KSE)*. IEEE, 1–7.
[11] Stuart Russell. 2022. Artificial Intelligence and the Problem of Control. *Perspectives on Digital Humanism* 19 (2022), 1–322.
[12] Dexter RR Scobee and S Shankar Sastry. 2019. Maximum likelihood constraint inference for inverse reinforcement learning. *arXiv preprint arXiv:1909.05477* (2019).