
Research Statement

Motivation

I seek to build intelligent agents that work for us and not against; i.e. it should contribute to the overall well-being of life rather than optimizing for short-term individualistic goal.

Research Interests: I am passionate about leveraging Deep Reinforcement Learning to design intelligent machines while building on the foundations of eXplainable AI (XAI) to improve user comprehension of these systems. Moreover, I perceive intelligence as a function of data as much as algorithms and view Knowledge Graphs as an integral part of an ideal intelligent system. I am also very open to working with the field of Cognitive Psychology and Public Policy to better position these intelligent systems.

My current research aims to address the question 'How can we leverage large datasets to build intelligent decision making engines that are robust, safe and explainable?'

Current Research

DeepAveragers and Extensions: Model Based Reinforcement Learning (MBRL) while very successful are yet to leverage optimal planners for decision making over search based planning methods. The high-level goal of this work is to move toward MBRL approaches that can effectively leverage near-optimal planners for improved data efficiency and flexibility in complex environments. We propose a very simple approach of combining deep representation with tabular models which can then be readily solved by using scaled value iteration. We find that the proposed approach scales to large datasets and more importantly allows us to use optimal planners for zero-shot transfer learning to changing environments and goals. We are currently working on extensions of the approach on continuous action spaces and further exploring use-cases such as optimal exploration and model compression/explainability.

User centric Explanation of Agents: As AI agents are getting more pervasive in daily life of users, it is important not only to have appropriate explanations but making sure that it is accessible to end users. To this end we study the need of users who have no AI/ML background when it comes to understanding when they should (or should not) trust the agent to act on their behalf? Moreover we also investigate how users react to the explanations in a game with a small action spaces?

Bias in Knowledge Graphs: Knowledge graphs are able to represent the same amount of information in more than one way; relating to structural variations of these graphs. As more machine learning algorithms are being applied directly on top of these graphs, we seek to investigate if these structural variations have any effect on downstream tasks. We find that these variations do affect the output of ML algorithms and try to quantify this further as a bias/variance trade-off among these representations. ¹

Looking Forward: I am not always the smartest person in the room, but I am mostly the most open minded. I feel it is now more important than ever to make AI research more inclusive and open without hindering the pace of its development. I look forward to being a part of a vibrant, diverse group reaching for the same goal.

Aayam K. Shrestha

📞 (541) 368 8894 • ✉ shrestaa@oregonstate.edu