Planning and Learning for Reliable Autonomy in the Open World

Sandhya Saisubramanian

Assistant Professor, Oregon State University, Oregon, USA

Safe and reliable decision-making is critical for long-term deployment of autonomous systems. Despite the recent advances in artificial intelligence, ensuring safe and reliable operation of human-aligned autonomous systems in open-world environments remains a challenge. My research focuses on developing planning and learning algorithms that support reliable autonomy in fully and partially observable environments, in the presence of uncertainty, limited information, and limited resources.

Summary of Recent Research

Due to the practical limitations in data collection and precise model specification, deployed systems often operate based on incomplete information, which results in undesirable behaviors. Model incompleteness may arise in the form of missing information that is critical to complete the assigned task (typically known unknowns), or missing superfluous details that do not affect the task performance but produce negative side effects (typically unknown unknowns). My group develops techniques to mitigate the undesirable impacts arising due to both types of model incompleteness. Some of our current research directions are described below.

1. Learning human-aligned reward functions: Agents often learn a *proxy* reward function when presented with expert demonstrations, since multiple reward functions may be consistent with the demonstrated behavior. To support safe deployment, it is necessary to ensure that the agent learns a reward function that is aligned with the demonstrator's intended reward. In a recent work (Mahmud, Saisubramanian, and Zilberstein 2022), we present an algorithm to reduce ambiguity associated with the learned reward, and validate reward alignment using explanations and verification tests.

2. Avoiding negative side effects: An autonomous system's actions may have negative side effects (NSEs), due to incomplete specification of objectives, constraints, or action effects. Avoiding NSEs involves the following steps: (1) gather information about NSEs from different forms of feedback; (2) learn a predictive model of NSEs to generalize the gathered information to unseen situations; and (3) plan to mitigate NSEs, without significantly affecting the task completion. We present learning and planning techniques to avoid *Markovian and non-Markovian* NSE (Srivastava et al. 2023; Saisubramanian, Kamar, and Zilberstein 2020, 2022). Markovian NSEs are learned and represented in a tabular format, and non-Markovian NSEs using a finite state machine. Planning using a learned NSE model is performed using a multi-objective approach with lexicographic reward preferences (**Distinguished Paper Award**, IJCAI 2020) (Saisubramanian, Kamar, and Zilberstein 2020), a human-agent team approach (Saisubramanian, Kamar, and Zilberstein 2022), and a constraint optimization approach (Srivastava et al. 2023).

3. Monitoring and restoring safety with minimal interference Deployed systems may perform unsafe actions when they encounter situations that are not fully described in their model. We use metareasoning to continuously monitor the underlying task process to detect safety violations and identify the right action to quickly recover from the situation, while minimally interfering with task completion (Svegliato et al. 2022). We are currently developing a metareasoning approach for safe operation of a team of cooperative agents.

Future Research Directions

In the near future, my group will continue to focus on reliable decision-making. In particular, we are developing techniques to enable autonomous systems to be cognizant of their limitations and biases, and adapt their behavior to overcome these limitations. We also aim to extend some of the above described contributions to partially observable and multi-agent settings.

References

Mahmud, S.; Saisubramanian, S.; and Zilberstein, S. 2022. REVEALE: Reward Verification and Learning Using Explanations. In *Submission*.

Saisubramanian, S.; Kamar, E.; and Zilberstein, S. 2020. A Multi-Objective Approach to Mitigate Negative Side Effects. In *Proc. of the 29th Intl. Joint Conf. on Artificial Intelligence*.

Saisubramanian, S.; Kamar, E.; and Zilberstein, S. 2022. Avoiding Negative Side Effects of Autonomous Systems in the Open World. *Journal of Artificial Intelligence Research*, 74: 143–177.

Srivastava, A.; Saisubramanian, S.; Paruchuri, P.; Kumar, A.; and Zilberstein, S. 2023. Planning and Learning for Non-Markovian Negative Side Effects Using Finite State Controllers. In *Proc. of the 37th AAAI Conference on Artificial Intelligence*.

Svegliato, J.; Basich, C.; Saisubramanian, S.; and Zilberstein, S. 2022. Metareasoning for Safe Decision Making in Autonomous Systems. In *Proc. of the Intl. Conf. on Robotics and Automation*.