Infinite Ends from Finite Samples: Open-Ended Goal Inference as Top-Down Bayesian Filtering of Bottom-Up Proposals

Tan Zhi-Xuan
Massachusetts Institute of Technology , Cambridge, Massachusetts, United States

Gloria Kang
Massachusetts Institute of Technology, Cambridge, Massachusetts, United States

Vikash Mansinghka
MIT, Cambridge, Massachusetts, United States

Josh Tenenbaum
MIT, Cambridge, Massachusetts, United States

Abstract

The space of human goals is tremendously vast; and yet, from just a few moments of watching a scene or reading a story, we seem to spontaneously infer a range of plausible motivations for the people and characters involved. What explains this remarkable capacity for intuiting other agents' goals, despite the infinitude of ends they might pursue? And how does this cohere with our understanding of other people as approximately rational agents? In this paper, we introduce a sequential Monte Carlo model of open-ended goal inference, which combines top-down Bayesian inverse planning with bottom-up sampling based on the statistics of co-occurring subgoals. By proposing goal hypotheses related to the subgoals achieved by an agent, our model rapidly generates plausible goals without exhaustive search, then filters out goals that would be irrational given the actions taken so far. We validate this model in a goal inference task called Block Words, where participants try to guess the word that someone is stacking out of lettered blocks. In comparison to both heuristic bottom-up guessing and exact Bayesian inference over hundreds of goals, our model better predicts the mean, variance, efficiency, and resource rationality of human goal inferences, achieving similar accuracy to the exact model at a fraction of the cognitive cost, while also explaining garden-path effects that arise from misleading bottom-up cues. Our experiments thus highlight the importance of uniting top-down and bottom-up models for explaining the speed, accuracy, and generality of human theory-of-mind.

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*