Domain Engineering to Represent Human Behavior Using Multi-Agent Planning and Inductive Methodologies

Salena Torres Ashton

School of Information, University of Arizona
1103 E. 2nd St. #409, Tucson, Arizona USA 85721
salena@arizona.edu

Abstract

This research combines multi-agent planning, the psycholinguistics of question-asking, procedural grounded theory, and hierarchical task networks to represent domains for automated planning.

Introduction

For my PhD research, I explore knowledge representation, coupled with machine learning and causal modeling, to infer the latent mental models of collaborating human problem solvers. Not only do I focus on model development, I actively research how assumptions inform knowledge engineering and data collection, which then influence model selection.

I designed an urban search and rescue (SAR) domain to model human planning during collaborative problem solving, which includes an explicit theory of mind model. We initially assumed teams of human players prioritized high-valued victims and navigation strategies to score more points. However, player behavior deviated from these hypotheses, so I proposed a new domain engineering strategy that accounts for all observed actions from each player, independent of any assumed strategy or inferred goal.

The methodology I developed combines four interdisciplinary research techniques:

1. **Multi-agent planning social protocols** (Baldoni et al. 2019) to represent player/agent coordination and commitment. Both the experiment design and this social protocol assume rational and normative agent behavior.
2. Event annotations, limited to **question-asking behaviors**. In this current phase of research, I focus on questions because they provide rich, asymmetrical information about collaborator belief and goals (Hawkins and Goodman 2017).
3. Inductive coding schema known as **Procedural Grounded Theory**, (GT), (Meireles et al. 2021) to annotate SAR videos.
4. Annotated events are represented as **Hierarchical Task Networks**, implemented in the Hierarchical Domain Definition Language (HDDL), then mapped to primitive actions and abstract tasks.

To my knowledge, this combination of methods is not found in current planning literature.

Background

A sufficiently complete domain hinges on clear problem definition, the skill and insight of domain engineers, and in specifying enough information about actions and tasks to account for all possible and relevant solutions (Bercher 2022; Gragera et al. 2023).

Domains are problematic when insufficient coverage of possible actions, goals, or solutions do not accurately model human planning behaviors. Such behaviors include: redundancy (repeated actions that add no obvious value), nuance (similar actions that have pragmatic significance), or inefficiency (adds to domain solution complexity).

Many modern planning models abstract away nuance, redundancy, and inefficiency, favoring technical approaches, empirical evaluations, and computational efficiency. However, these approaches are less accurate at capturing subtle and relevant human behavior (Barberis Canonico, McNeese, and Duncan 2018). They could introduce bias, especially when models are not informed by theories of human behavior (Kang 2023; Pereira, Pereira, and Meneguzzi 2021; Sreedharan et al. 2020; Weber and Bryce 2011) that use nuance or redundancy as input. Representing inefficient human behavior with technically efficient approaches could also introduce an ethical risk when output from less-interpretable models are implemented in real-world scenarios like SAR, impacting real people.

Current Research

To emulate human planning behavior within our domains, we can use *mixed* methodologies such as "human in the loop" (Barberis Canonico, McNeese, and Duncan 2018), mitigating ethical risk. **What key domain engineering features might we overlook when we disregard nuance in human behavior?** For example, humans can *request* help, allowing the response to be either a 'yes' or 'no', or they *ask* for help, expecting the answer to be 'yes' because a 'no' would be impolite. This kind of subtlety relies upon the important role of pragmatics in human communication.

Data and Methods

To represent human nuance and pragmatics in planning domains, I used Procedural Grounded Theory (GT) and Baldoni's social protocol framework for multi-agent coordination as the conceptual framework of our label development. Three fully-autonomous annotators, all from our plan recognition work group, annotated 12 SAR videos, clarifying labels and analyzing question-asking behaviors with their inferred goals. GT is known to be an initially expensive coding process for the first few observations, yet this cost monotonically decreases with each new observation. This mixed method took 13:30 man hours, combined, for three annotators and twelve observations. I then rigorously analyzed the labels, using axial and selective GT to discern the data-driven patterns and emergent theories of team planning behavior.

Results and Conclusions

We achieved an unweighted Cohen's Kappa score of 0.892 for evaluations, demonstrating reliability and reproducibility. Using this method for observing and representing human planning behavior, I found that:

(1) Teams who ask more questions at a *consistent* rate scored higher than teams who *sporadically* asked the same number of questions. (2) Compared to all forms of natural language communication, teams who asked more questions scored more points. (3) All teams demonstrated redundancy, nuance, and inefficiency. However, teams who asked questions to *clarify* an idea, compared to teams who instructed each other, showed increased collaboration, rescued more victims, and scored higher points. (4) Teams who coordinated on tasks that did not require collaboration, compared to teams who collaborated only when required, scored more points and exhibited increased question-asking behaviors.

These results provided the framework design of the HTN structures for primitive actions and abstract tasks, which I then implemented in HDDL. These HTNs account for planning strategies that include high score prioritization and navigation strategies, as well as unexpected actions like repeating questions, team plan deviation, and inefficient team work.

HTN structures and domains crafted with this method are replicable, domain-agnostic, and explainable, ensuring their intelligibility across academic disciplines and practical real-world applications.

Contributions and Future Work

While complex models may reduce computational complexity, they may not authentically reflect inefficient human behaviors. My mixed method for domain design generates information that naturally guides the construction of HTNs to represent human planning behavior, which is easily implemented in HDDL. This work is currently developing a unique dataset that relates actions, abstract tasks, goals, and natural language. The next phase of this research would include model comparisons and empirical evaluation metrics.

My dissertation research combines this method, formal semantics, and text extraction of causal and counterfactual language. After completing my PhD, I will continue to promote interdisciplinary research effort between artificial intelligence, social history research, and causal modeling.

Acknowledgments

Thank you to my reviewers for their encouraging feedback, and to Adarsh Pyarelal, Loren Reiffer-Champlin, and Clayton T. Morrison for their support and guidance. This research was supported by DARPA's Artificial Social Intelligence for Successful Teams (ASIST) program.

References

Baldoni, M.; Baroglio, C.; Micalizio, R.; and Tedeschi, S. 2019. Programming Agents by Their Social Relationships: A Commitment-Based Approach. *Algorithms*, 12(4): 76.

Barberis Canonico, L.; McNeese, N. J.; and Duncan, C. 2018. Machine Learning as Grounded Theory: Human-Centered Interfaces for Social Network Research through Artificial Intelligence. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 62(1): 1252–1256.

Bercher, P. 2022. Hierarchical Planning and Reasoning about Partially Ordered Plans — From Theory to Practice. *AI Magazine*, 43(4): 353–364.

Gragera, A.; Fuentetaja, R.; Garcia-Olaya, A.; and Fernández, F. 2023. Repair Suggestions for Planning Domains with Missing Actions Effects. *Proceedings of the International Conference on Automated Planning and Scheduling*, 33(1): 153–161.

Hawkins, R.; and Goodman, N. D. 2017. Why do you ask? The informational dynamics of questions and answers. Published on the Open Science Foundation website, PsyArXiv preprint. Accessed: 2023-06-23.

Kang, E. B. 2023. Ground truth tracings (GTT): On the epistemic limits of machine learning. *Big Data & Society*, 10(1): 205395172211461.

Meireles, A.; Carvalho, R. M.; Rique, T.; Cavalcante, M. B. P.; Perkusich, M.; Almeida, H.; and Perkusich, A. 2021. Towards a Grounded Theory for a Development Process Model for Machine Learning Based Systems. In *Anais do I Workshop Brasileiro de Engenharia de Software Inteligente (ISE 2021)*, 19–24. Sociedade Brasileira de Computação.

Pereira, R. F.; Pereira, A. G.; and Meneguzzi, F. 2021. Landmark-Enhanced Heuristics for Goal Recognition in Incomplete Domain Models. *Proceedings of the International Conference on Automated Planning and Scheduling*, 29(1): 329–337.

Sreedharan, S.; Chakraborti, T.; Muise, C.; Khazaeni, Y.; and Kambhampati, S. 2020. A Case Study of XAIP in a Model Acquisition Task for Dialogue Planning. *Proceedings of the International Conference on Automated Planning and Scheduling*, 30: 488–497.

Weber, C.; and Bryce, D. 2011. Planning and Acting in Incomplete Domains. *Proceedings of the International Conference on Automated Planning and Scheduling*, 21: 274–281.