TA-endorsed List (Grace Song)

Advancing Social Goal Inference: Integrating Social Choice Theory in Bayesian Multi Agent Models

Name:Chance Jiaje Li (graduate)
Email: jiajie@mit.edu

In this research proposal, I aim to enhance the domain of social goal inference within multiagent systems by integrating social choice theory into Bayesian models. This endeavour is intriguing as it extends the capabilities of AI agents to consider group preferences and collective welfare in their decision-making processes, a significant step towards sophisticated, socially-aware AI. The research question I seek to address is: How can the integration of social choice theory into Bayesian models improve the inference of social goals in multiagent settings, particularly in scenarios involving collective actions and resource negotiation?

To build a solid foundation for this research, I will draw upon the following recent and relevant works:

1. Reimplement and Extend Bayesian Models: I will reimplement the Bayesian model of social goal inference as outlined by Ullman et al. (2009), and extend it to include social choice dynamics.
2. Design Multi Agent Interaction Scenarios: Create scenarios where agents engage in voting and negotiation processes, reflecting the principles of social choice theory.
3. Collect and Analyze Experimental Data: Conduct experiments to collect data on these interactions and analyse the data to assess the efficacy of the integrated model.
4. Model Comparison and Evaluation: Compare the outcomes with traditional models to quantify improvements in social goal inference.

References:

1. *Ullman, T. D., Baker, C. L., Macindoe, O., Evans, O., Goodman, N. D., & Tenenbaum, J. B. (2009). Help or Hinder: Bayesian Models of Social Goal Inference. Advances in Neural Information Processing Systems, 22.*
2. *Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2011). Bayesian theory of mind: Modelling joint belief-desire attribution. Proceedings of the Thirty-Third Annual Conference of the Cognitive Science Society.*
3. *Jern, A., & Kemp, C. (2015). A decision network account of reasoning about other people's choices. Cognition, 142, 12-38.*
4. *Lucas, C. G., Griffiths, T. L., Williams, J. J., & Kalish, M. L. (2015). A rational model of function learning. Psychonomic Bulletin & Review, 22(5), 1193-1215.*
5. *Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. Trends in Cognitive Sciences, 20(11), 818-829.*