NRI: FND: Scalable and Customizable Intent Inference and Motion Planning for Socially-Adept Autonomous Vehicles

Wenlong Zhang (PI), Yi Ren (Co-PI), and Yezhou Yang (Co-PI), Arizona State University

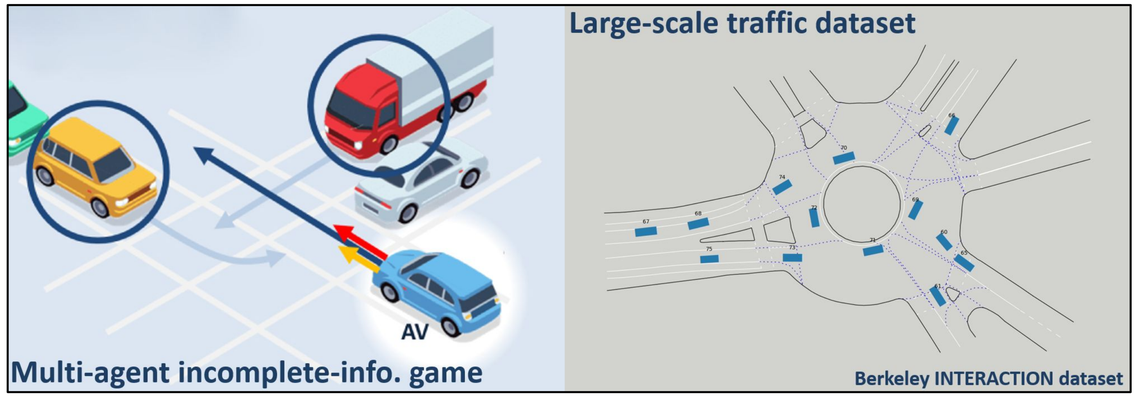
https://home.riselab.info/nri.html

Objectives: To improve the scalability, generalizability, robustness, and social adeptness of real-time intent inference and motion planning algorithms in collaborative robots (co-robots), with an application in autonomous vehicles.

Challenges:

- Interactions are differential games w/ incomplete-information
- Computing perfect Bayesian equilibrium (PBE) in real time in computationally expensive
- Real-world applications require building efficient multi-modal systems for the edge-device.

Solutions: Theory- and data-driven PBE approximation, a detector-free and end-to-end Vision Transformer based one-stage multi-modal model.



Scientific Impact:

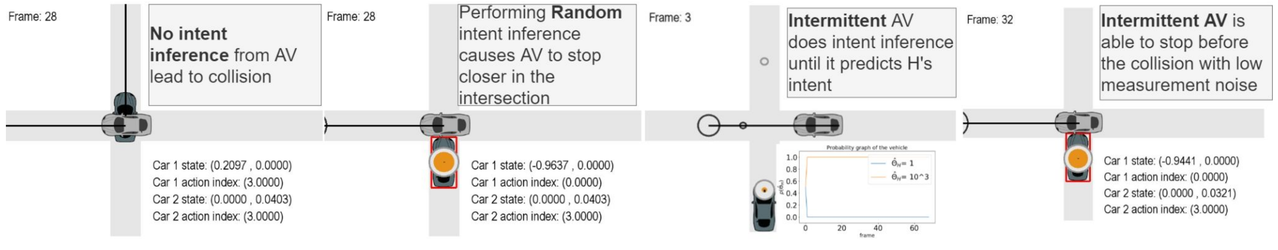
- Accurate intent inference and effective signaling during interactions
- Safe and human-like intent inference via event-driven intent inference model.

Broader Impact:

- Collaboration with Arizona Institute of Automated Mobility/Intel (ongoing)
- 2 REU interns on conf. papers and honor thesis (ongoing)
- Interdisciplinary course on human-robot interaction (in plan)

Result 1: Intermittent Empathetic Intent Inference [1]

- RL agent decides when to perform intent inference and thus is able to significantly reduce the computational costs and maintain safe interaction - Video
- The proposed method is able to predict correct intent even in the presence of noise in the measurement - Video



Result 2: Empathetic Intent Inference with Belief Uncertainty [2]

- Inference via bounded rationality, Bayesian belief update & equilibrial Hamiltonian approximation
- Hamiltonian incorporate belief uncertainty, approximated via learning of co-state dynamics
- Empathy improves intent inference accuracy, and interaction safety in turn, especially when agents have biased prior beliefs (e.g., not recognizing extremely aggressive drivers) - Video

Result 3: Efforts towards Building Efficient Multi-modal System [3]

- To understand the complex traffic scenarios, building a reliable and efficient AI system is necessary that jointly encodes inputs from different *modalities*.
- We develop an efficient multi-modal system that connects the inputs from Vision and Language, and is not reliant on any object detector so is end-to-end trainable.
- Compared with the previous methods, our model reduces 90% of the inference GFLOPs and reaches SOTA performances on the image captioning benchmark.

1. *Amatya, S., Ghimire, M., Ren, Y., Xu, Zhe., Zhang, W., "When Shall I Estimate Your Intent? Costs and Benefits of Intent Inference in Multi-Agent Interactions", ACC2022 (accepted)*
2. *Chen, Y., Zhang, L., Merry, T., Amatya, S., Zhang, W., Ren, Y., "When Shall I be Empathetic? The Utility of Empathetic Parameter Estimation in Multi-Agent Interactions", ICRA2021*
3. *Fang, Z., Wang, J., Hu, X., Liang, L., Gan, Z., Wang, L., Yang, Y. and Liu, Z., 2021. "Injecting Semantic Concepts into End-to-End Image Captioning." CVPR 2022*

2022 NRI & FRR Principal Investigators' Meeting
April 19-21, 2022

Award ID#: 1925403