Learning communication in action models for human-robot collaboration

Contact:

Mohamed Chetouani
ISIR - CNRS UMR7222 Sorbonne Université
mohamed.chetouani@sorbonne-universite.fr

Context:

When humans demonstrate tasks, they integrate instrumental actions (manipulating objects) with ostensive communicative cues (e.g., eye gaze, pauses, and exaggeration). These belief-directed actions are intentional, aimed at conveying information beyond the immediate task. This concept, termed communication in action, has been widely studied in Cognitive Science as a key mechanism for establishing shared understanding between actors and observers [Ho-2022].

In robotics, communication in action (Figure 1) is mirrored by the challenge of generating legible robot motion [Dragan-2013,Wallkötter-2022]. Legible motions combine instrumental actions (achieving goals) with belief-directed cues to make the robot's intent clear to human observers. This often involves exaggerated movements or gaze cues, enabling humans to infer the robot's goal accurately [Wallkötter-2021]. Computational approaches in Human-Robot Interaction (HRI) typically rely on mathematical models of human expectations to design such motions. However, most current approaches assume a literal interpretation of human and robot actors, leading to limitations in understanding and responding to each other's actions and intentions. These limitations highlight the need for forward and inverse models capable of encoding and decoding both instrumental and belief-directed intentions, enabling mutual comprehension in human-robot collaborations.

OSTENSIVE is a recent ANR project starting in 2025, coordinated by ISIR and involving CNRS LAAS and INRIA LARSEN. This project aims to advance HRI by developing human-centric interaction systems that integrate instrumental and belief-directed dimensions by exploiting recent machine learning approaches. A PhD project is planned to be proposed as a follow-up to the internship.

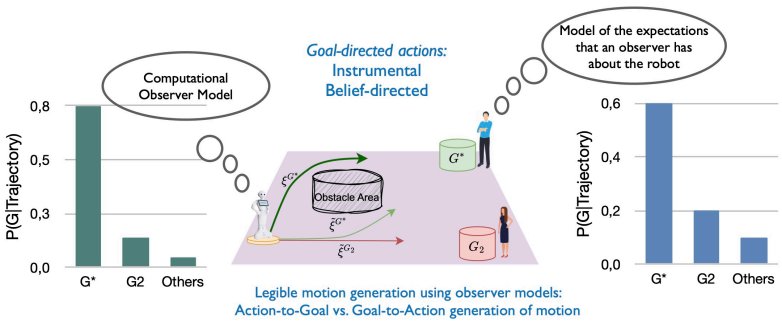

Figure 1 Communication in action for HRI. Legible motion generation is a typical goal-to-action situation using mathematical observer models of humans allowing to perform instrumental and belief-directed actions.

Objectives:

This internship focuses on the development of machine learning models that enable robots to act based on double goals: achieving tasks (instrumental actions) and communicating intent (belief-directed actions). For this purpose, we will exploit behavior cloning approaches based on variational auto-encoders and/or diffusion policies [Lee-2024; Liang-2024; Caselles-2022].

The specific objectives are:

- Develop forward and inverse models to encode and decode instrumental and belief-directed actions.
- Design machine learning algorithms to integrate these dimensions into robot behavior.
- Test these models in tasks where robots are instructed through combined instrumental and belief-based goals.

Profile:

Level: Master 2 / Engineer school
Skills: Python, Machine Learning, Robotics, and Cognitive Science
Duration: 5-6 months

References:

[CasellesDupré-2022] Caselles-Dupré, H., Sigaud, O., Chetouani, M. (2022) Pragmatically Learning from Pedagogical Demonstrations in Multi-Goal Environments, NeuRIPS 2022.

[Dragan-2013] Dragan, A, Lee, Srinivasa, S. (2013) Legibility and predictability of robot motion, IEEE/ACM HRI.

[Ho-2022] Ho M. and Griffiths T. L. (2022) Cognitive science as a source of forward and inverse models of human decisions for robotics and control, Annual Review of Control, Robotics, and Autonomous Systems, 5, 33-53.

[Liang-2024} Zhixuan Liang, Yao Mu, Hengbo Ma, Masayoshi Tomizuka, Mingyu Ding, Ping Luo; SkillDiffuser: Interpretable Hierarchical Planning via Skill Abstractions in Diffusion-Based Task Execution Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024, pp. 16467-16476

[Lee-2024} Lee, S., Wang, Y., Etukuru, H., Kim, H.J., Shafiullah, N.M.M. & Pinto, L.. (2024). Behavior Generation with Latent Actions. Proceedings of the 41st International Conference on Machine Learning.

[Wallkötter-2021] Wallkötter, S., Tulli, S., Castellano, G., Paiva, A., Chetouani, M. (2021-. Explainable Embodied Agents Through Social Cues: A Review. J. Hum.-Robot Interact. 10, 3, Article 27.

[Wallkötter-2022] Wallkötter S., Chetouani M., Castellano G. (2022). SLOT-V: Supervised Learning of Observer Models for Legible Robot Motion Planning in Manipulation, IEEE RO-MAN 2022.