

Sujet de thèse

Titre de la thèse : MODÈLES COMPUTATIONNELS DE PRISE DE PERSPECTIVE POUR L'INTERACTION HUMAIN-ROBOT

Directrice ou directeur de thèse : Mohamed Chetouani

Co-direction éventuelle : Malika Auvray

Laboratoire d'accueil : ISIR (*Institut des Systèmes Intelligents et de Robotique*), Campus Pierre et Marie Curie, 4 place Jussieu, 75005 Paris.

Personne à contacter

Prénom Nom : Mohamed Chetouani

Tel : +33 1 44 27 63 08

Email : mohamed.chetouani@sorbonne-universite.fr

Envoyer votre candidature par mail, avec [sujet de la thèse] en objet, un CV et une lettre de motivation.

Date limite de dépôt de la candidature : 22 Mai 2022

Description du sujet (en français)

Contexte et Objectifs :

L'aptitude à prendre la perspective d'autrui pendant une interaction contribue significativement à notre capacité à interagir. Cette thèse a pour objectif le développement de modèles computationnels permettant à un robot de prendre en compte le point de vue centré humain pour la génération d'expressions en langage naturel et de mouvements pendant des tâches collaboratives.

La plupart des robots collaboratifs seront mis œuvre dans un environnement peuplé d'humains où ils devront accomplir des missions. Ils seront engagés en interaction avec des humains concernés par les mêmes missions. La compréhension mutuelle et la coordination des objectifs, des intentions, des plans et des actions sont nécessaires pour réaliser ces missions. Durant ces interactions, l'humain et le robot échangent via des un ensemble de signaux verbaux et non-verbaux.

Dans cette thèse, nous nous focalisons sur deux importantes modalités de communication à savoir le langage et le mouvement. L'objectif de cette thèse est de permettre à un robot de prendre le point de vue de l'humain partenaire afin de générer des expressions en langage naturel et des mouvements adéquats. Les modèles computationnels doivent permettre de reproduire les aptitudes de changement de perspective afin de décrire une scène à l'aide du langage naturel (e.g., relations spatiales des objets) et générer des mouvements compréhensibles par l'humain.

Les processus de prise de perspective sont étudiés dans plusieurs domaines de recherche. Plusieurs travaux en sciences cognitives ont en particulier montré la diversité des dimensions dans la prise de perspective telles que les dimensions visuo-spatiales, cognitives ou encore émotionnelles [Schilbach-2013, Hamilton-2014]. Les humains ont des préférences individuelles

Sous la co-tutelle de :

de prise de perspective, en particulier la flexibilité à changer de perspective [Arnold-2016].

En interaction humain-robot, l'aptitude de prise de perspective d'un robot est critique pour le succès de tâches collaboratives. Dans [Dogan-2020], un modèle computationnel permet à un robot de décrire les relations spatiales entre les objets en adoptant le point de vue de l'humain (e.g., "la bouteille que vous voyez à droite de la banane"). Le modèle exploite des techniques d'apprentissage pour modéliser les relations spatiales en fonction de la position de l'humain (e.g., pour un robot placé en face de l'humain, la bouteille est à gauche de la banane). Concernant la génération de mouvements, la plupart des travaux portent sur la génération de mouvements dits "lisibles" ("legible"), qui permettent à un robot de communiquer son intention à un observateur humain [Dragan-2013]. En s'inspirant de la recherche sur la façon dont les humains interprètent les comportements observés, les approches computationnelles visent à maximiser l'inférence humaine d'un objectif étant donné les mouvements du robot. On parle alors de transparence par le mouvement, cette aptitude est d'autant plus importante lors de l'utilisation de robots tels que les bras manipulateurs. Les travaux de cette thèse se situent dans cette lignée, mais visent à y intégrer le point de vue de l'humain, qui jusque-là était soit ignoré.

Description du projet :

Cette thèse exploite des méthodologies issues de l'interaction humain-robot, de l'apprentissage automatique et des sciences cognitives pour développer et évaluer des modèles computationnels de prise de perspective pour la génération d'expressions en langage naturel et la génération de mouvements.

Les travaux de cette thèse s'organisent autour de deux axes de recherche complémentaires :

(1) Modélisation des préférences et des capacités de prise de perspectives de l'humain en interaction avec un robot

Les travaux en sciences cognitives ont fourni plusieurs approches expérimentales pour l'évaluation pour des préférences et des capacités de prise de perspective chez l'humain. Cependant, ces méthodes ne sont pas directement utilisables en interaction humain-robot. En effet, l'apparence du robot (humanoïde ou pas) et/ou les mouvements générés (mouvement de tête, geste de pointage) semblent influencer sur les capacités du partenaire humain à changer de perspective [Zhao-2022].

Cet axe a pour objectif de développer des méthodologies et des modèles adaptés à l'interaction humain-robot. Dans cet objectif, il est judicieux d'exploiter l'interaction entre l'humain et le robot en proposant une tâche à même d'évaluer les compétences de l'humain. On pourra en particulier s'intéresser à l'imitation et l'apprentissage par démonstration dans des tâches de manipulation d'objets. Il s'agira de développer des modèles d'apprentissage (en particulier des auto-encodeurs) permettant de représenter le monde en fonction des préférences et des capacités de prise de perspective de l'humain.

(2) Génération d'expressions en langage naturel et de mouvements par un robot

Cet axe a pour objectif la génération d'expressions en langage naturel et de mouvements par le robot en exploitant les modèles de prise de perspective. Il s'agira, en particulier, de combiner des modèles d'apprentissage de modélisation de relation (e.g., relational networks) avec des modèles de raisonnement Bayésien. Cette approche exploite la représentation des relations spatiales des objets par apprentissage. Le formalisme du raisonnement Bayésien permet la flexibilité de prise de perspective (égocentrique, allocentrique). Ces modèles seront utilisés pour interagir avec l'humain en utilisant les relations entre les objets pour le dialogue et la manipulation d'objets.

Sous la co-tutelle de :

Les modèles développés dans ces deux axes seront évalués dans des tâches collaboratives avec des métriques quantitatives (temps de réponse, décision) et qualitatives (questionnaires d'expérience utilisateur).

Profil recherché et compétences requises :

- Diplômé(e) de Master 2 / École d'Ingénieur (ou équivalent) en Robotique, Intelligence Artificielle ou Sciences Cognitives
- Expérience souhaitée en robotique expérimentale (stage)
- Compétences en informatique (Python, C++)
- Bonne maîtrise des méthodes d'apprentissage automatique
- Grande motivation pour les travaux et approches centrés humain
- Bonne maîtrise de l'anglais

Références :

[Arnold-2016] Arnold, G., Spence, C., & Auvray, M. (2016). Taking someone else's spatial perspective: Natural stance or effortful decentring? *Cognition*, 148, 27-33.

[Dogan-2020] Dogan, F., Gillet, S., Carter, E. and Leite, I. (2020) Journal Article, *Robotics and Autonomous Systems*, Vol. 134, December.

[Dragan-2013] Dragan, AD, Lee, KCT, Srinivasa, SS. (2013) Legibility and predictability of robot motion, 8th ACM/IEEE Inter. Conf. on Human-Robot Interaction (HRI 2013)

[Hamilton-2014] Hamilton A., Kessler K., Creem-Regehr S. (2014) Perspective taking: building a neurocognitive framework for integrating the "social" and the "spatial". *Front. Hum. Neurosci.*

[Ho-2021] Ho M. and Griffiths T. L. (2021) 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* 2022 5:1

[Job-2022] Job, X.E., Kirsch, L.P. & Auvray, M. Spatial perspective-taking: insights from sensory impairments. *Exp Brain Res* 240, 27–37 (2022).

[Sebanz-2006] Sebanz, N, Bekkering, H, Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences*, 10, 70-76.

[Schilbach-2013] Schilbach, L, Timmermans, B, Reddy, V, Costall, A, Bente, G, Schlicht, T, Vogeley, K (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*.

[Trafleton-2005] Trafleton, J., Gregory et al. "Perspective-taking with robots: experiments and models." *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication*, 2005.(2005): 580-584.

[von Mohr-2020] Von Mohr M, Finotti G, Ambroziak KB, Tsakiris M. (2020) Do you hear what I see? An audio-visual paradigm to assess emotional egocentricity bias. *Cogn Emot.* 34, 756-770.

[Wallkotter-2021] Wallkotter S., Tulli S., Castellano G., Paiva A., Chetouani M. (2021). Explainable Embodied Agents Through Social Cues: A Review, *ACM Trans. on Hum-Robot Interact.* 10, 3, Article 27

Sous la co-tutelle de :

[Wallkotter-2022] Wallkotter, S., Chetouani, M., Castellano, G. (2022) A new approach to evaluating legibility: Comparing legibility frameworks using framework independent robot motion trajectories, arXiv:2201.05765

[Zhao-2022] Zhao, X., Malle, B. F. (2022) Spontaneous perspective taking toward robots: The unique impact of humanlike appearance, Cognition, Volume 224, 105076

Description du sujet (en anglais)

Context and objectives:

The ability to take the perspective of others during an interaction contributes significantly to our ability to interact. The objective of this thesis is to develop computational models that allow a robot to take into account the human-centered point of view for the generation of natural language expressions and movements during collaborative tasks.

Most of the collaborative robots will be implemented in an environment populated by humans where they will have to accomplish missions. They will be engaged in interaction with humans involved in the same missions. Mutual understanding and coordination of goals, intentions, plans and actions are required to accomplish these missions. During these interactions, the human and the robot exchange via a set of verbal and non-verbal signals.

In this thesis, we focus on two important communication modalities, namely language and motion. The objective of this thesis is to allow a robot to take the point of view of the human partner in order to generate natural language expressions and appropriate movements. The computational models must be able to reproduce the skills of perspective change in order to describe a scene using natural language (e.g., spatial relationships of objects) and generate human-understandable movements.

Perspective taking processes are studied in several research areas. In particular, several works in cognitive science have shown the diversity of dimensions in perspective taking such as visuo-spatial, cognitive or even emotional dimensions [Schilbach-2013, Hamilton-2014]. Humans have individual preferences for perspective taking, in particular the flexibility to change perspective [Arnold-2016].

In human-robot interaction, a robot's perspective-taking ability is critical for the success of collaborative tasks. In [Dogan-2020], a computational model allows a robot to describe spatial relationships between objects by adopting the human's perspective (e.g., "the bottle you see to the right of the banana"). The model exploits learning techniques to model the spatial relations according to the position of the human (e.g., for a robot placed in front of the human, the bottle is to the left of the banana). Concerning motion generation, most of the work focuses on the generation of so-called "legible" motions, which allow a robot to communicate its intention to a human observer [Dragan-2013]. Drawing on research on how humans interpret observed behaviors, computational approaches aim to maximize human inference of a goal given the robot's movements. This is called motion transparency and is especially important when using robots such as manipulator arms. The work of this thesis is in this line, but aims at integrating the human point of view, which until now was either ignored.

Project Description:

This thesis exploits methodologies from human-robot interaction, machine learning and cognitive sciences to develop and evaluate computational perspective-taking models for natural language expression generation and motion generation.

Sous la co-tutelle de :

The work in this thesis is organized around two complementary research directions:

(1) Modeling the preferences and perspective-taking abilities of humans interacting with a robot.

Work in cognitive science has provided several experimental approaches for the evaluation of human preferences and perspective-taking abilities. However, these methods are not directly applicable to human-robot interaction. Indeed, the appearance of the robot (humanoid or not) and/or the movements generated (head movement, pointing gesture) seem to influence the human partner's abilities to change perspective [Zhao-2022].

This axis aims at developing methodologies and models adapted to human-robot interaction. To this end, it is wise to exploit the interaction between the human and the robot by proposing a task that can evaluate the human's skills. In particular, we will be able to focus on imitation and learning by demonstration in object manipulation tasks. The aim is to develop learning models (in particular auto-encoders) that can represent the world according to the preferences and perspective-taking abilities of the human.

(2) Generation of natural language expressions and movements by a robot

The objective of this axis is to generate natural language expressions and movements by the robot by exploiting the perspective taking models. In particular, it will combine learning models of relationship modeling (e.g., relational networks) with Bayesian reasoning models. This approach exploits the representation of spatial relations of objects by learning. The Bayesian reasoning formalism allows for flexibility in perspective taking (egocentric, allocentric). These models will be used to interact with humans by using the relations between objects for dialogue and object manipulation.

The models developed in these two axes will be evaluated in collaborative tasks with quantitative (response time, decision) and qualitative (user experience questionnaires) metrics.

Required Profile and skills:

- Master 2 / Engineering school: Robotics, Artificial Intelligence or Cognitive Science
- Experience in experimental robotics (optional)
- Programming skills (Python, C++)
- Previous knowledge in machine learning
- High motivation of human centered works and approaches
- Very good English skills

Sous la co-tutelle de :