TITRE DE LA THESE:
RAPALLO: Robotic Action Planning with Affordances and Large Language mOdels

Direction de thèse : Panagiotis PAPADAKIS, Ehsan ABBASNEJAD

Co-encadrant·es : Mihai ANDRIES

Laboratoire(s) :

☐ GEPEA ☐ IRISA ☒ Lab-STICC ☐ LATIM
☐ Lego ☐ LEMNA ☐ LS2N ☐ hors Laboratoire

Equipe(s) de recherche : RAMBO

Département(s) IMT Atlantique :

☐ DAPI ☐ DSEE ☒ INFO ☐ ITI ☐ LCI ☐ LUSSI
☐ MEE ☐ MO ☐ OPT ☐ SSG ☐ SRCD ☐ SUBATECH

S'agit-il d'une thèse en cotutelle internationale ?
☒ **Oui** ☐ **Non**
Si oui, organisme avec lequel la cotutelle est envisagée :
University of Adelaide, Adelaide, Australia

Le sujet proposé présente-il un caractère interdisciplinaire ?
☒ **Oui** ☐ **Non**
Si oui, expliquer brièvement pourquoi (2 ou 3 lignes) :
Le sujet est un mélange de robotique et d'apprentissage par renforcement.

La source du co-financement est-elle identifiée ?
☒ **Oui** ☐ **Non**
Si oui, préciser quel co-financement est envisagé :
Co-financement du côté de University of Adelaide.

Autres informations :
Informations utiles que vous souhaiteriez communiquer (si pertinent) :

Keywords: robotics, action planning, reinforcement learning, affordance, large language models

1 Context / State-of-the-art

Robotic action planning is the process that allows a robot to plan a sequence of actions leading to a desired goal state. Traditionally, this is done using tools like the PDDL language, which allows to define a so-called *planning domain* that describes the possible actions in each state, as well as their preconditions and effects. The challenge with such an approach is how to populate the domain of possible actions that a robot can execute. In autonomous robots, this planning domain has to be independently constructed by discovering the possible actions in the environment. Existing research efforts focus on building affordance maps, that require recognizing which interactions are possible with objects in the environment. Such methods are based on prior knowledge linking object properties with robotic actions and observed effects. Despite numerous attempts [1, 5, 4], autonomous affordance detection and action planning is still an open problem.

An effective way to enhance a robot's ability to reason and plan about the world is to incorporate a world model. With the recent advancements in the language [2] and multimodal pre-trained models [6] trained on large web corpora, there is a promising opportunity to use them to aid in reasoning [3] and planning. By leveraging these models, robots can generate alternative plans based on various samples obtained from them, depending on the context, simulating the imagination of the agent. This approach also enables robots to perform open-ended tasks without the need for prior training on a specific problem.

2 Objectives of the thesis

Expected social and economic impact: This thesis will be a stepping stone towards the development of action planning capabilities of autonomous robots. If successful, it will allow robots to learn how to interact with objects in the environment with much less training data, as compared to standard reinforcement learning approaches.
Scientific question: Can large language models improve robotic action planning based on affordances? In addition, how to overcome the computation challenge when using large models?
Outcomes of the PhD: (1) Feasibility analysis of using large pre-trained language and vision-and-language models in robotics to evaluate affordances in the world; (2) A robotic action-planning model that utilises large pre-trained language and vision-and-language models for planning and reasoning about the environment.

3 Competences expected from potential candidates

- Master Degree in Computer Science (or equivalent)
- Programming and Software Engineering skills (Python)
- Machine learning skills, in particular Reinforcement Learning and Bayesian Networks
- Deep Learning skills (PyTorch)

References

[1] Alper Ahmetoglu, M Yunus Seker, Justus Piater, Erhan Oztop, and Emre Ugur. "DeepSym: Deep Symbol Generation and Rule Learning for Planning from Unsupervised Robot Interaction". In: *Journal of Artificial Intelligence Research* 75 (2022), pp. 709–745.

[2] Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. "Language models are few-shot learners". In: *Advances in neural information processing systems*. Vol. 33. 2020, pp. 1877–1901.

[3] William Chen, Siyi Hu, Rajat Talak, and Luca Carlone. "Leveraging Large Language Models for Robot 3D Scene Understanding". In: *arXiv preprint arXiv:2209.05629* (2022).

[4] Alexander Khazatsky, Ashvin Nair, Daniel Jing, and Sergey Levine. "What can i do here? learning new skills by imagining visual affordances". In: *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2021, pp. 14291–14297.

[5] George Konidaris, Leslie Pack Kaelbling, and Tomas Lozano-Perez. "From skills to symbols: Learning symbolic representations for abstract high-level planning". In: *Journal of Artificial Intelligence Research* 61 (2018), pp. 215–289.

[6] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. "CLIP: Contrastive Language-Image Pre-training". In: *arXiv preprint arXiv:2103.00020* (2021).