Structured representation learning using geometrical groups for human-inspired AI

This is an interdisciplinary project for an intern who will work on representation learning structured by geometrical groups, with a strong focus on implementation. The trainee will work on theoretical questions, implementations and evaluation strategies.

Background

Recent advances in computational psychology and artificial agents modelling have leveraged principles of group theory to model information integration and action planning based on intrinsic and extrinsic reward (Rudrauf et al., 2022, see also Rudrauf et al., 2020; Rudrauf et al., in press). Agents have a world model, encoding spatial and valuative information, which they access and represent through the action of the 3D projective group as a method for perspective taking. They use group action to anticipate maximally informative and rewarding actions, and perform active inference, for non-social and social planning.

Aims

The approach requires at this point a pre-defined world model. We are now seeking to integrate geometrically structured representation learning in our approach so that agents can learn how to send new observations into their world model. More specifically, in recent work (Sergeant et al. 2022), we proved that group structured representations of world models can have a big impact on exploration strategies. We now want to go a step further and have agents learn exploration strategies in simulated environments based on limited information on their environment using group structured latent representations. In order to do so, we need to train architectures that learn to represent observations in a latent space that is geometrically structured. We will want to test the extent to which geometry can thus contribute to the quality of the learnt representation and to quantifying epistemic value, for adaptive exploration behaviors.

Outstanding issues

The project entails algorithmic challenges, e.g. how to perform back propagation when the latent space is geometrically structured by groups. It implies to integrate concept from geometry, information theory and inference into optimal algorithms.

Perspectives

We ultimately seek to develop solutions for autonomous virtual and robotic agents, to model and simulate human-like cognition and behavior, but also toward the implementation of human-inspired, human-oriented and interpretable artificial systems, with a wealth of scientific and industrial applications.

Requirements for the intern

- basics notions in group theory (definitions)
- basics in computer science, including experience with python and basic knowledge of pytorch

Co-supervision

The project will be co-supervised by:

- Grégoire Sergeant-Perthuis, postdoctoral scholar in mathematics, INRIA team Ouragan & Sorbonne University mathematics departement team combinatorics and optimization (IMJ-PRG)
- Yaël Frégier, Enseignant-Chercheur (MCU) in pure mathematics and machine learning, Artois University
- David Rudrauf, Professor (CPJ) in cognitive science, CIAMS Laboratory, University Paris-Saclay

About the team:

It is a pluridisciplinary team, with a wide experience of working together, a strong background in mathematics and machine learning (deep learning, reinforcement learning) and cognitive science, offering a rich and unique supervision.

Contact : david.rudrauf@universite-paris-saclay.fr

References

Rudrauf, D., Sergeant-Perthuis, G., Belli, O., Tisserand, Y., & Serugendo, G. D. M. (2022). Modeling the subjective perspective of consciousness and its role in the control of behaviours. Journal of Theoretical Biology, 534, 110957.

Rudrauf, D., Sergeant-Perthuis, G., Tisserand, Y., Monnor, T., & Belli, O. (in press). Combining the Projective Consciousness Model and Virtual Humans to assess ToM capacity in Virtual Reality: a proof-of-concept. ACM TiiS. See: arXiv preprint arXiv:2104.07053.

Rudrauf, David, Daniel Bennequin, and Kenneth Williford. "The moon illusion explained by the projective consciousness model." Journal of theoretical biology 507 (2020): 110455.

Sergeant-Perthuis, G, Rudrauf, D, Tisserand,Y. '(2022) Curiosity driven exploration through perspective transformation'. https://www.researchgate.net/publication/366262449_Curiosity_driven_exploration_through_perspective_transformation