Presentation: Sparsity Regularisation for Sample Efficient Learning of Symbolic Planning Operators

Henri-Jacques Geiss[1,*], Alejandro Agostini[1], Ingmar Posner[2] and Justus Piater[1]

[1]*Department of Computer Science, University of Innsbruck, Austria*
[2]*Applied AI Lab, University of Oxford, UK*

Abstract
In line with Image Schemas being natural groundings of symbolic representation, we investigate how autonomous agents can learn generalizable symbolic predicates of properties of objects and their relations that can be used for long-horizon sequential decision making. As the basis for our approach serves the Relational DeepSym framework, that learns a symbolic planning domain solely based on random exploration with time-extended skills in a multi-object robotic workspace. Here, we demonstrate that directly regularising for increased sparsity of the learned binary symbols improves sample efficiency and therefore leads to better planning performance in a set of object-manipulation tasks. Additionally, we compare symbolic planning with continuous baselines for forward prediction and discuss the interpretability of the derived symbols and operators.

Keywords
Robot Learning, Symbol Emergence, Sparsity Regularisation, PDDL

1. Introduction

In her theory on "How to build a baby" [1], Jean Mandler defines Image Schemas as spatially structured representations of the relations between entities and their movement through space that constitute *earliest meanings* and a *natural grounding of symbolic representation* in human cognition. To date, it is still an open question whether the bridge between human perception and higher level cognitive abilities, like conceptual thinking and language acquisition, is formed by image schematic conceptual primitives. Regarding *artificial* agents, recent work has presented strategies how Image Schemas can be leveraged for hierarchical decision making, especially based on their symbolic nature [2, 3, 4].

For instance, by using a translation of Image Schema Logic to the Web Ontology Language, Pomarlan et al. [2] synthesized an "Image Schematic Reasoning Layer" that infers queries about the state of the environment to be answered by a separate perception module. This way perception can be performed in a focused manner, by only attending to entities that are part of the emitted queries and therefore considered relevant in the current state. Such approaches predominantly rely on predefined and handcrafted predicates such as *contained* or *inFrontOf* for their symbolic rules.

To that end, the "DeepSym" line of research [5, 6, 7] by Ahmetoglu et al. established an approach for bottom-up symbol learning based on sensorimotor exploration of a robot in a workspace with multiple objects. These symbols can then subsequently be used for generating a set of operators that capture the effect of actions in the symbolic space in the Planning Domain Definition Language (PDDL), which in turn enables long-horizon sequential decision making with of-the-shelf planners [8]. In its most recent version, the proposed encoding architecture for symbol learning also includes relational symbols between pairs of objects. On one hand, the explicit learning of relational symbols showcases the potential of "Relational DeepSym" [6] (RDS) to serve as an architecture for Image-Schema-like symbol emergence in artificial agents. On the other hand, the RDS framework does not contain any

The 9th Image Schema Day, International Conference on Formal Ontology in Information Systems (FOIS) and the Joint Ontology Workshops Episode XI (JOWO), September 8-12 2025, Catania, Italy
*Corresponding and presenting author.
henri.geiss@uibk.ac.at (H. Geiss); alejandro.agostini@uibk.ac.at (A. Agostini); ingmar@robots.ox.ac.uk (I. Posner); justus.piater@uibk.ac.at (J. Piater)
0000-0001-6270-700X (I. Posner); 0000-0002-1898-3362 (J. Piater)

inductive bias towards learning explicitly *generalizable* predicates that capture large parts of the state space and would enable successful long-horizon planning from observing a small amount of transitions for the PDDL domain generation.

2. Method

In this presentation, to meet this desideratum in autonomous embodied symbol acquisition, we investigate how *sparsity regularisation* influences the learned predicates and their quality for learning a PDDL planning domain from randomly collected transitions.

Vectors, e.g. of parameters, encodings or atttention masks, are sparser the more of their elements are close to or at zero [9]. Lei et al. [10] recently demonstrated how to learn a causal world model by applying sparsity regularisation to a transformer with hard attention. Given that the unary and relational predicates that are learned in the RDS architecture are binary by design, sparsity in this case can be enforced by simple L_0 regularisation on the learned symbolic encoding.

To evaluate the benefits of learning sparse symbolic representations for action planning we present a simulated robot workspace environment, that contains multiple objects of different types and a set of time-extended skills like *Pick* or *MoveOver* to manipulate them. Based on this environment, we probe the effect of sparsity on the generalizability of the symbols with a set of test tasks consisting of planning problems with 2 to 4 objects in the workspace. We find that due to the increased sparsity, larger regions of the state space are captured by each symbol, leading to a decrease in the amount of observed transitions necessary for a certain performance on the planning tasks.

The sparsest relation between two objects that can be learned is that they do not influence each other at all and the symbolic relation vector is only zeros. For that reason, as found in prior work [10], applying sparsity regularisation to the RDS architecture also leads to learning when objects are explicitly *not* related i.e. have no effect on each other. We lay out how this can be used for further improvements in data efficiency in the PDDL domain generation. Since planning in symbolic space is not an end in itself, we additionally compare our approaches' performance with several continuous baselines.

Lastly, we discuss the interpretability of the learned symbols and operators. For that, we present where, and where not, learned predicates align with conventional Image Schemas and compare the inferred PDDL Domain to one that a human expert would create to solve the given set of tasks.

In summary, we present the results of our investigation of how to artificially mimic image schematic natural grounding of symbolic representation by applying sparsity regularisation to the RDS architecture for planning symbol learning. We find that sparsity does improve sample efficiency in the PDDL domain generation and forces the explicit encoding of the absence of any relation between two objects, yet, general alignment with human symbols and planning domains is still limited and offers anchor points for future work.

References

[1] J. M. Mandler, How to build a baby: Ii. conceptual primitives., Psychological review 99 (1992) 587.

[2] M. Pomarlan, S. De Giorgis, M. M. Hedblom, M. Diab, N. Tsiogkas, Thinking in front of the box: Towards intelligent robotic action selection for navigation in complex environments using image-schematic reasoning., in: The Joint Ontology Workshops, Location: Jönköping, Sweden, 2022.

[3] M. Pomarlan, S. De Giorgis, R. Ringe, M. M. Hedblom, N. Tsiogkas, Hanging around: Cognitive inspired reasoning for reactive robotics, in: Formal Ontology in Information Systems, IOS Press, 2024, pp. 2–15.

[4] N. Tsiogkas, Using image schemata to support autonomous assembly tasks., in: ISD, 2022.

[5] A. Ahmetoglu, M. Y. Seker, J. Piater, E. Oztop, E. Ugur, Deepsym: Deep symbol generation and

rule learning for planning from unsupervised robot interaction, Journal of Artificial Intelligence Research 75 (2022) 709–745.
[6] A. Ahmetoglu, B. Celik, E. Oztop, E. Ugur, Discovering predictive relational object symbols with symbolic attentive layers, IEEE Robotics and Automation Letters 9 (2024) 1977–1984.
[7] A. Ahmetoglu, E. Oztop, E. Ugur, Symbolic manipulation planning with discovered object and relational predicates, IEEE Robotics and Automation Letters (2025).
[8] M. Helmert, The fast downward planning system, Journal of Artificial Intelligence Research 26 (2006) 191–246.
[9] E. Mathieu, T. Rainforth, N. Siddharth, Y. W. Teh, Disentangling disentanglement in variational autoencoders, in: International conference on machine learning, PMLR, 2019, pp. 4402–4412.
[10] A. Lei, B. Schölkopf, I. Posner, Spartan: A sparse transformer learning local causation, arXiv preprint arXiv:2411.06890 (2024).