DECSTR: LEARNING GOAL-DIRECTED ABSTRACT BEHAVIORS USING PRE-VERBAL SPATIAL PREDICATES IN INTRINSICALLY MOTIVATED AGENTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Past research showed that preverbal infants represent abstract spatial relations between physical objects and demonstrate goal-oriented sensorimotor behavior before they even start acquiring language. Inspired by these findings, we introduce DECSTR, an intrinsically motivated agent endowed with a high-level representation of spatial relations between objects that freely explores its environment and sets its own goals as target constraints on these spatial relations. In a robotic manipulation environment, DECSTR explores the corresponding representation space by manipulating objects, and efficiently learns to achieve any reachable configuration within it. It does so by leveraging an object-centered modular architecture, a symmetry inductive bias, and a new form of automatic curriculum learning for goal selection and policy learning. We explicitly compare the use of this abstract state and goal representations to the traditional goal-as-state framework and show that DECSTR can fulfill goals in an opportunistic manner, making profit of the current configuration of objects to achieve its goals. Additionally, by contrast with traditional instruction-following agents which are trained simultaneously to understand language and to act, DECSTR decouples goal-oriented sensorimotor learning from language acquisition, as in infants. This decoupling is facilitated by the abstract spatial relations, as they better match the natural semantics of linguistic inputs than traditional continuous representations. We show this behavior in a proof-of-concept setup, where trained agents learn to associate their behavioral repertoire with simple language commands via a language-conditioned goal generator. We show that this goal generator increases the behavioral diversity of a given language instruction and enables retry behaviors when the agent fails.

1 INTRODUCTION

Developmental psychology investigates the interactions between learning and developmental processes that support the slow but extraordinary transition from the mental life of infants to the sophisticated intelligence of human adults (Piaget, 1977; Smith & Gasser, 2005). Inspired by this line of thought, the central endeavour of developmental robotics consists in shaping a set of machine learning processes able to generate a similar growth of capabilities in robots (Weng et al., 2001; Lungarella et al., 2003). This approach is particularly interested in the *Symbol Grounding Problem* (Harnad, 1990), the challenge of explaining how a machine can extract the meaning of symbols by grounding them into non-symbolic sensorimotor interactions (Sigaud & Droniou, 2016).

The design of intrinsically motivated agents marked a major step in developmental robotics. These embodied agents interact with their environment at the sensorimotor level and are endowed with the ability to represent and set their own goals, rewarding themselves over completion (Forestier et al., 2017). They often come with the capacity to automatically organize their learning curriculum (Portelas et al., 2020), deciding which goals to target and learn about as a function of their current abilities (Florensa et al., 2017; Fournier et al., 2019; Colas et al., 2019a; Racaniere et al., 2019).

Natural language, because of its capacity of abstraction, has been recently connected to sensorimotor behaviors by the reinforcement learning (RL) community. On the one hand, *language-conditioned* agents have been introduced (Chan et al., 2019; Bahdanau et al., 2018; Cideron et al., 2019; Jiang

et al., 2019; Luketina et al., 2019; Colas et al., 2020). However, all these approaches use language as a necessary input to sensorimotor behaviors and, for this reason, cannot account for the goal-directed behaviors observed in pre-verbal infants (Wood et al., 1976). On the other hand, *language grounding* has emerged to map natural instructions to play (Cangelosi et al., 2010; Lynch & Sermanet, 2020; Tellex et al., 2011; Alomari et al., 2017). Some of these approaches further integrate a symbolic intermediate representation to handle diverse language produced by untrained users (Tellex et al., 2011). However, all these works use either teleoperated agents (Lynch & Sermanet, 2020; Alomari et al., 2017) or labeled hand-engineered skills (Tellex et al., 2011).

In order for instrinsically motivated agents to learn sensorimotor skills before mastering language, many algorithms were introduced within the goal-directed RL framework (Schaul et al., 2015; Andrychowicz et al., 2017). In these works, the goal space is defined as a subset of the state space of the agents (Nair et al., 2018; Florensa et al., 2018; Pong et al., 2019; Nair et al., 2019). This does not provide an adequate level of abstraction to learn and communicate with other artificial or human agents about these goals, which may be necessary for the kind of symbolic intelligence found in children (Bruner, 2009; Taniguchi et al., 2018). Indeed, developmental psychology has shown that pre-verbal infants are equipped with a set of conceptual relational spatial primitives (Mandler, 2012). These forms of predicates over object relations enable pre-verbal abstract thoughts and, later on, support early language acquisition (Mandler, 2012).

Based on these findings, and to circumvent the limitations of the direct mapping between language and behavior that is ubiquitous in instruction following agents, we propose DECSTR[1]. DECSTR is a learning architecture training intrinsically motivated agents to explore a semantic representation space characterizing spatial relations between physical objects. From access to object-related features and a few spatial relation predicates, DECSTR learns to discover and master all reachable configurations in its semantic representation space. Then, in a proof of concept setup, we show that this language-agnostic semantic representation facilitates language acquisition when coupled to a language-conditioned goal generator[2].

To summarize, this paper introduces the following contributions: 1) the concept of leveraging pre-verbal semantic representations for the acquisition of abstract behavioral repertoires; 2) a new approach to language grounding that can be decoupled from sensorimotor learning. These conceptual contributions are supported by technical contributions: 1) new object-centered inductive biases; 2) a novel automatic curriculum approach and 3) a language-conditioned goal generation module. This study is conducted in the *Fetch Manipulate* environment, where the DECSTR agent is embodied in a simulated robotic arm facing blocks, and must learn to manipulate them to discover and master various configurations from the semantic representation space.

2 RELATED WORK

From language-conditioned policies to predicate-based policies. Language was recently introduced as a way to help or instruct goal-conditioned learning agents (Luketina et al., 2019). Most of these works consider *instruction-following agents* (Hermann et al., 2017; Chan et al., 2019; Bahdanau et al., 2018; Cideron et al., 2019; Jiang et al., 2019; Fu et al., 2019), whereas the Imagine approach introduced intrinsically motivated agents able to set their own goals and to imagine new ones (Colas et al., 2020). Other approaches used descriptive language to build representations (Waytowich et al., 2019) or to characterize the dynamics of the environment (Zhong et al., 2019). Our approach is different: because the policy is not conditioned on language, we can decouple language grounding from the acquisition of sensorimotor skills.

Symbol grounding using relational representations Many works investigate the use of symbolic relational representations to associate meanings to robotic behaviors (Alomari et al., 2017; Tellex et al., 2011; Kulick et al., 2013). On the one hand, some of these works consider these symbolic representations as an intermediate layer between language and sensorimotor skills (Alomari et al., 2017; Tellex et al., 2011). Indeed, Alomari et al. (2017) explore learning language groundings using video clips by automatically clustering behaviors into linguistic inputs using relational graph representations. Tellex et al. (2011) propose a method that uses symbolic meaning representations

[1]for "DEep sets and Curriculum with SemanTic goal Representations".

[2]Code and videos can be found at `https://sites.google.com/view/decstr/`.

based on Generalized Grounding Graphs (G^3). On the other hand, Kulick et al. (2013) address the acquisition of symbolic representations without leveraging natural language. They propose an active learning scenario to map robotic behaviors to pre-existing symbols. All these works use input skills that are either manually controlled (Alomari et al., 2017), hand-engineered (Tellex et al., 2011) or obtained with explicit optimal control methods such as trajectory optimization (Kulick et al., 2013). In contrast, our intermediate abstract representation is used to train our agent to learn the sensorimotor skills and to ground language.

Blocks manipulation Stacking blocks is one of the earliest benchmarks in artificial intelligence (e.g. Sussman (1973); Tate et al. (1975)) and has led to many simulation and robotics studies, e.g. (Deisenroth et al., 2011; Xu et al., 2018; Colas et al., 2019a). Like us, several recent papers have addressed complex block manipulation in the Fetch OpenAI environment (Plappert et al., 2018). Lanier et al. (2019) achieve stacks of 4 blocks, whereas Li et al. (2019) achieve a variety of goals with up to 6 blocks. Both works represent goals as a set of specific 3D target positions for each block and rely on hand-defined curriculum. Indeed, in Li et al. (2019), additional blocks are added to the experiment once the performance of the agent passes a specified threshold. In Lanier et al. (2019), stages with different complexities are considered, where the generated target positions and the starting blocks configuration depend on the current stage. In contrast, since DECSTR uses semantic goal representations, it can decide on the locations of the blocks as long as the spatial relations described by the semantic goal are verified. Besides, it learns its own curriculum by biasing goal sampling towards sets of goals associated to high-learning progress (Colas et al., 2019a; Fournier et al., 2019). These sets of goals are gradually constructed according to the time of discovery.

Relational inductive biases Previous work has advocated for the use of inductive biases in learning architectures (Battaglia et al., 2018). In particular, relational inductive biases can help learning interactions with objects characterized by their relations. Colas et al. (2020) first presented an object-centered modular RL architecture to improve transfer of skills between objects, which was then extended to consider object pairs in reward prediction (Karch et al., 2020). Here, we leverage the symmetry properties of some spatial relations as additional inductive biases.

3 METHODS

We consider agents equipped with innate semantic representations of their states, evolving in an unknown environment containing physical objects. These semantic representations can be formalized as a list of spatial relations between objects in the scene. We propose the DECSTR learning architecture. First, during *sensorimotor learning*, the agent leverages an automatic curriculum mechanism to sample from the set of discovered goals, boostrapping from random configurations. It alternates between goal-directed interactions with objects and actual learning updates. It updates both the list of known goals and the sampling probabilities of its goal sampler, as well as its policy using the Soft Actor-Critic algorithm (Haarnoja et al., 2018) and Hindsight Experience Replay (Andrychowicz et al., 2017). Then, to show the efficiency of our semantic representations, we introduce a *proof of concept* setup where DECSTR receives simple language descriptions of its goal-directed behavior and uses them to train a language-conditioned goal generation module that links language to skills, effectively grounding language in sensorimotor behavior.

In this section, we first introduce the *Fetch Manipulate* environment designed to study the acquisition of goal-directed sensorimotor behaviors based o abstract semantic representations (Section 3.1). Then, we highlight the technical contributions of the DECSTR algorithm: new forms of object-centered inductive biases (Section 3.2) and a novel automatic curriculum strategy (Section 3.3). Finally, we present our proof of concept setup where we use a novel language-conditioned goal generator which enables language grounding (Section 3.4).

3.1 ABSTRACT SEMANTIC REPRESENTATIONS IN THE *Fetch Manipulate* ENVIRONMENT

Semantic configurations. We assume that agents have access to innate representations called *semantic configurations*. These configurations are based on a list of predicates describing spatial relations between pairs of objects in the scene. In this paper, we consider two of the spatial predicates infants demonstrate early in their development (Mandler, 2012): the *close* and the *above* binary

predicates. These two predicates are applied to all permutations of object pairs, i.e. 6 permutations for the 3 objects we consider. Because the *close* predicate is order invariant, we only need to evaluate it on 3 object combinations. The *above* predicate being order dependent, we need all 6 permutations. The resulting binary vector of size 9 forms the *semantic configuration*. It represents the spatial relations between objects in the scene. In the resulting semantic configuration space $\{0, 1\}^9$, the agent can reach 35 physically valid configurations, including stacks of 2 or 3 blocks and pyramids. Note that, in principle, the DECSTR learning architecture could use any other combination of binary predicates and could be extended to use n-ary predicates. Appendix A provides formal definitions and properties of predicates and semantic configurations.

The *Fetch Manipulate* environment. To evaluate DECSTR, we introduce the *Fetch Manipulate* environment: a robotic manipulation domain based on MUJOCO (Todorov et al., 2012) and derived from the Fetch tasks (Plappert et al., 2018). Actions are 4-dimensional where 3 dimensions specify the movement in Cartesian coordinates and the last dimension controls the gripper. Observations include features of the gripper and the considered objects. Since we manipulate 3 objects, goals are defined as 9-dimensional semantic configurations. The binary reward function directly derives from the semantic mapping: the agent rewards itself when its current configuration c_p matches the goal configuration $c_p = g$. Appendix B provides examples of valid configurations.

3.2 Object-Centered Inductive Biases

DECSTR leverages two types of object-centered inductive biases in the design of its policy and critic. The first inductive bias lies in the design of the networks themselves. A shared network independently encodes body features and each pair of objects features, as well as the current and target configurations. This is inspired from architectures presented in Colas et al. (2020); Karch et al. (2020), based on Deep Sets (Zaheer et al., 2017). This shared network ensures efficient transfer of skills between object pairs. The second inductive bias leverages the symmetry of the behavior required to achieve *above*$(o_i,\ o_j)$ and *above*$(o_j,\ o_i)$. To ensure automatic transfer between the two, we present half of the features (e.g. those based on pairs $(o_i,\ o_j)$ where $i < j$) with goals containing one side of the symmetry (all *above*$(o_i,\ o_j)$ for $i < j$) and the other half of the features with the goals containing the other side of the symmetry (all *above*$(o_j,\ o_i)$ for $i < j$). As a result, the *above*$(o_i,\ o_j)$ predicates fall into the same slot of the shared network inputs as their symmetric counterparts *above*$(o_j,\ o_i)$, only with different permutations of object pairs. The goals are now of size 6: 3 *close* and 3 *above* predicates, corresponding to one side of the *above* symmetry. Skill transfer between symmetric predicates are automatically ensured. Appendix C.2 further describes these inductive biases and our modular architecture.

3.3 Curriculum Learning

DECSTR agents use an automatic curriculum strategy (Portelas et al., 2020) inspired from the CURIOUS algorithm (Colas et al., 2019a). They track aggregated estimations of their *competence* (C), *learning progress* (LP). They bias their selection of goals to target during data collection and goals to learn about during policy updates towards goals associated with high absolute LP and low C.

Automatic bucket generation. To facilitate robust estimation, LP is usually estimated on sets of goals with similar difficulty or similar dynamics (Forestier et al., 2017; Colas et al., 2019a). While previous works leveraged expert-defined *goal buckets*, we propose a strategy that groups goals together based on their time of discovery. The intuition is that the time of discovery is a good proxy for goal difficulty: easy goals are discovered early, more difficult goals later. Agents start with no known configurations (empty buckets). When they end an episode in a new configuration, the $N_b = 5$ buckets are updated. Buckets are filled equally and the first buckets contain the configurations discovered earlier. For this reason, goals change buckets as new goals are discovered.

Tracking competence, learning progress and sampling probabilities. Regularly, agents perform self-evaluations: they evaluate themselves on goal configurations sampled uniformly from the set of known ones. For each bucket, the agent tracks the recent history of past successes and failures when targeting the corresponding goals (last $W = 1800$ self-evaluations). C is estimated as the success rate over the most recent half of that history $\text{C} = \text{C}_{\text{recent}}$. LP is estimated as the difference

between the current C (C_{recent}) and the one evaluated over the first half of the history ($C_{earlier}$). This is a crude estimation of the derivative of the C curve w.r.t. time: LP = C_{recent} - $C_{earlier}$. The sampling probability P_i for bucket i is then computed as:

$$P_i = \frac{(1 - C_i) * |LP_i|}{\sum_j ((1 - C_j) * |LP_j|)}.$$

In addition to the usual LP bias (Colas et al., 2019a), this formula favors lower C when LP estimates are fairly similar. The absolute value ensures a regain of interest for buckets that show decreased performances (e.g. forgetting). Implementation details are provided in Appendix C.

Help from a social partner. After a first period of discovering and learning to reach configurations, agents are helped by sporadic interventions of a social partner. Helping the agents in their exploration, the social partner can organize the scene with non-trivial initial configurations (see details in Appendix B). This is reminiscent of the dynamics of parent-infant interactions as characterized by the concept of Zone of Proximal Development (Vygotsky, 1978).

3.4 Language Module

We introduce a language-conditioned goal generation module that generates semantic configurations matching the agent's initial configuration and a sentence describing an expected transformation of one object-pair relation. We call this language-conditioned generative model the *language module* thereafter. This section explains how it is trained and used for language grounding.

Language-conditioned goal generation We implement the language module with a conditional Variational Auto-Encoder (C-VAE) (Sohn et al., 2015). A training dataset is collected via interactions between a trained DECSTR agent and a social partner. For each goal-directed trajectory the agent performs, the social partner provides the description of one of the resulting transformations in the object relations. The set of possible descriptions contains 102 sentences, each describing, in a simplified language, a positive or negative shift for each of the 9 predicates (e.g. *get red above green*). The C-VAE is similar to the one proposed in Nair et al. (2019). We simply add an extra condition on the instruction that describes the expected transformation. This way, we add a control on goal generation. This instruction is encoded via a recurrent network that is jointly trained. Appendix C.4 provides the list of sentences and implementation details.

Language grounding At test time, DECSTR is instructed to modify an object relation by one of the 102 sentences. Conditioned on its current configuration and instruction, it samples a compatible goal from the language module. This module effectively enables agents to ground NL in their internal semantic representations and set of sensorimotor skills. We consider three evaluation settings: 1) performing a single instruction; 2) performing a sequence of instructions; 3) performing a logical combination of instructions. As the agent can generate a set of goals matching any instruction, it can easily combine these sets to perform logical functions of instructions: *and* is an intersection, *or* is an union and *not* is the complement within the set of goals the agent discovered during sensorimotor training. Given sets of compatible goal configurations, agents can also *try again*: find other target configurations that match the required instruction when previous attempts failed.

4 Experiments

We first present experiments to evaluate the performance of sensorimotor learning (Phase 1), before considering language grounding (Phase 2). Learning curves report medians and interquartile ranges of the average success rates across goals ($\overline{SR}$). These are computed over 5 seeds for baselines and ablations and 10 seeds for DECSTR and its oracle variant with expert-defined buckets. Stars indicate significant differences w.r.t. the DECSTR condition, as reported by Welch's t-tests with $\alpha = 0.05$ (Colas et al., 2019b).

4.1 SENSORIMOTOR LEARNING PHASE: DISCOVERING AND MASTERING SEMANTIC GOAL CONFIGURATIONS

DECSTR. DECSTR successfully masters all reachable configurations in its semantic representation space. Figure 1a shows the evolution of $\overline{SR}$ computed per bucket. The fact that these buckets are learned in increasing order confirms that the time of discovery is a good proxy for difficulty. Figure 1b reports C, LP and sampling probabilities P as computed online for an agent using self-evaluations. The agent leverages these estimations to select its goals: first focusing on the easy goals from bucket 1, it moves on towards harder and harder buckets as easier ones are mastered (low LP, high C). Appendix E presents more examples of learning trajectories, and dissects the evolution of bucket compositions along training.

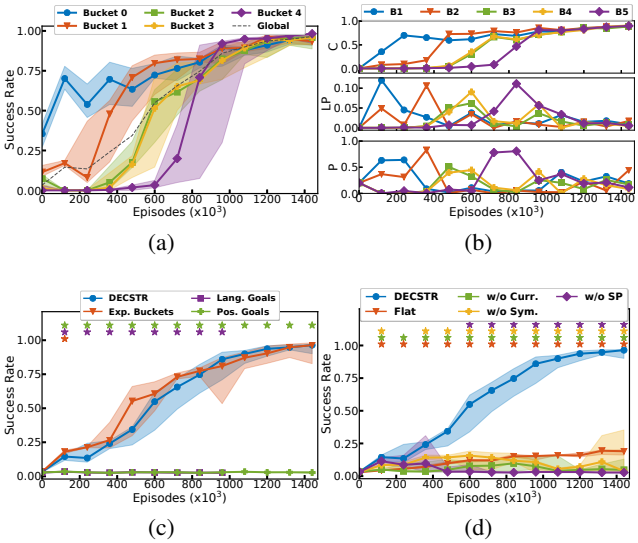

Figure 1: **Sensorimotor Learning**: (a) $\overline{SR}$ per bucket. (b): C, LP and P as estimated online by a DECSTR agent. (c) Baselines comparisons. Note that the *Language Goals* condition stops at 950×10^3 episodes for a same computing budget (3 days). (d) Ablations.

Baselines. We compare DECSTR to three different baselines. All of them are variants of DECSTR, use the same goal configuration selection based on curriculum learning, and are evaluated with the same metric: matching between goal configuration and final configuration. The *Expert Buckets* baseline uses an expert similarity-based definition of buckets and, as such, can be considered an oracle w.r.t. the automatic curriculum component of DECSTR. It also leverages a prior knowledge of the set of reachable configurations. The *Language Goals* and *Position Goals* baselines do not have access to their current configuration or goal configuration. Instead, the *Language Goals* baseline uses a jointly learned embedding of a sentence corresponding to that goal configuration. The *Position Goals* samples a 3D position target for each block so that the resulting semantic configuration matches the semantic goal (see Appendix D for extended definitions). Figure 1c shows that DECSTR performs on par with the *Expert Buckets* condition and dramatically improves over the *Language Goals* and *Position Goals* baselines. We hypothesize that the low performance of the *Language Goals* condition is due to 1) the difficulty to link the continuous current state to the abstract goal expressed through language and 2) by the drifting of the goal representation during training. In the *Position Goals* condition, we presume low performance is due to the sparsity of the task, as agents need to precisely position each block at their respective targets.

Opportunistic goal reaching. When targeting a given semantic goal configuration, the *Position Goals* baseline samples particular target positions for each block. This enforces a total distance d_{pos} for the blocks to travel from initial to target configurations. Because it does not use position targets, we hypothesized that DECSTR would find a shorter path d_{decstr}. Computed over 50 episodes for each of the 35 valid goals, we found that $\Delta = d_{pos} - d_{decstr}$ is positive 96% of the time. The

average across repetitions is superior to 19 cm for all goals, with a mean of 27 cm (all $p < 10^{-10}$). This highlights an advantage of using semantic goal configuration over 3D position targets, as done in most other architectures (Lanier et al., 2019; Li et al., 2019).

Ablation studies. Figure 1d presents the results of our ablation studies. Each condition removes one component of DECSTR: 1) *Flat* replaces our object-centered modular architectures by flat ones; 2) *w/o Curr.* replaces our automatic curriculum strategy by a uniform goal selection; 3) *w/o Sym.* does not use the symmetry inductive bias; 4) *w/o SP* the social partner does not provide non-trivial initial configurations. Only the full version of our architecture enables agents to discover and learn to master all reachable configurations.

4.2 Language Grounding Phase

In the *sensorimotor* learning phase, DECSTR agents acquire a repertoire of abstract skills without access to language. In the *language grounding phase*, we first train a language-conditioned goal generation module from data collected via interactions between DECSTR and a social partner (Section 3.4). For a given initial configuration and a given sentence, we want the language module to generate all compatible final configurations, and just these.

Language-conditioned goal generation performance. To evaluate the language module, we construct a synthetic, oracle dataset $\mathcal{O}$ of triplets $(c_i, s, \mathcal{C}_f(c_i, s))$, where c_i is the initial configuration, s is the sentence describing the expected transformation and $\mathcal{C}_f(c_i, s)$ is the set of all final configurations compatible with (c_i, s). Note that, on average, $\mathcal{C}_f$ in $\mathcal{O}$ contains 16.7 configurations, while the training dataset $\mathcal{D}$ only contains 3.4 (20%). We are interested in two metrics: 1) The *Compatibility Probability* (CP) is the probability that a goal sampled from the generator belongs to $\mathcal{C}_f$; 2) The *Coverage* (Cov) is the size of the intersection between $\mathcal{C}_f$ and the set resulting from sampling the generator 100 times. We compute these metrics on 5 different sets of input pairs (c_i, s), each calling for a different type of generalization:

1. Pairs found in $\mathcal{D}$, except pairs removed to form the following test sets. This calls for the extrapolation of known initialization-effect pairs (c_i, s) to new final configurations c_f ($\mathcal{D}$ contains only 20% of $\mathcal{C}_f$ on average).
2. Pairs that were removed from $\mathcal{D}$, calling for a recombination of known effects s on known c_i.
3. Pairs for which the c_i was entirely removed from $\mathcal{D}$. This calls for the transfer of known effects s on unknown c_i.
4. Pairs for which the s was entirely removed from $\mathcal{D}$. This calls for generalization in the language space, to generalize unknown effects s from related sentences and transpose this to known c_i.
5. Pairs for which both the c_i and the s were entirely removed from $\mathcal{D}$. This calls for the generalizations 3 and 4 combined.

Our language module demonstrates these 5 types of generalization (see Table 1). Agents can generate goals from situations they never encountered (3). They can generalize the meaning of sentences they never heard (4) and even apply the latter to unknown situations (5). We provide the content of testing sets in Appendix C.4. We trained the equivalent module to generate continuous target positions for all blocks, which is how the *Position Goals* condition represents goals. Result show strongly decreased performance, see Appendix F.

Table 1: Language module average metrics over 10 seeds. Std is below 0.07 for Cov and 0.06 for CP.

Metr.	Test 1	Test 2	Test 3	Test 4	Test 5
CP	0.93	0.94	0.95	0.90	0.92
Cov	0.97	0.93	0.98	0.99	0.98

Table 2: Language grounding performance metrics over 10 seeds (av±std).

Metr.	Transition	Expression
SR1	0.89 ± 0.05	0.74 ± 0.08
SR5	0.99 ± 0.01	0.94 ± 0.06

Grounding language in sensorimotor behavior. We investigate how the language module interacts with the sensorimotor skills of the agent. We consider three evaluation settings. In the *transition* setup, we monitor the average success rate of the agent when asked to perform the 102 instructions 5 times each, resetting the environment each time. In the *expression* setup, we evaluate the agent on 500 randomly generated combinations of elementary sentences. In both setups, we give the agent

5 attempts, enabling it to resample new compatible goals when the previous failed (without reset). Success rates after 1 (SR1) and 5 (SR5) attempts are reported in Table 2. In the *sequence* setup, the agent must execute 20 random sequences of instructions without reset. We report the average number of successes before the agent fails: $L = 14.9 \pm 5.7$ (std). Our 10 DECSTR agents are evaluated with the 10 C-VAE models evaluated above. These results highlight efficient language grounding. Agents achieve instructed transitions almost all the time, resampling alternative goals when previous ones failed, which only happens when blocks were previously kicked out of reach.

5 DISCUSSION AND CONCLUSION

This paper introduces DECSTR, a learning architecture that discovers and masters all reachable configurations from a set of relational spatial primitives, before undertaking an efficient language grounding phase. This was made possible by the use of object-centered inductive biases, a new form of automatic curriculum learning and a novel language-conditioned goal generation module.

Semantic representations Results have shown that using predicate-based representations was sufficient for DECSTR to efficiently learn abstract goals in an opportunistic manner. We believe that the proposed semantic configurations showcase some promising properties: 1) they reduce the complexity of the block stacking benchmarks where most effective works rely on a heavy hand-crafted curriculum (Li et al., 2019; Lanier et al., 2019) and a curiosity mechanism (Li et al., 2019), 2) they facilitate the grounding of language into sensorimotor behaviors as our proof of concept study has shown and 3) they enable decoupling sensorimotor learning from language grounding, as observed in infants (Piaget, 1977).

A new approach to language grounding. The language grounding approach proposed here as a proof of concept is the first to enable a decoupling between sensorimotor learning and language grounding for policy learning agents. In addition, it fosters a diversity of possible behaviors for any given instruction. Indeed, while an instruction-following agent trained on goals like *put red close_to green* would just push the red block towards the green one, our agent can generate many matching goal configurations. It could build a pyramid, make a blue-green-red pile or target a dozen other compatible configurations. This enables agents to *try again*, to find alternative approaches to satisfy a same instruction when first attempts failed. Our goal generation module can also generalize to new sentences or transpose instructed transformations to unknown initial configurations. Finally, the goal generation module automatically enables agents to deal with any logical expression made of instructions by combining generated goal sets. Note that nothing in the architecture prevents language from being grounded during sensorimotor learning, which would result in "overlapping waves" of sensorimotor and linguistic development (Siegler, 1998).

Semantic configurations of variable size. Considering a constant number of blocks and, thus, fixed-size configuration spaces is a current limit of our approach. We could imagine scaling to inputs of variable sizes by leveraging Graph Neural Networks as in Li et al. (2019). Corresponding semantic configurations could be represented as a set of vectors, each encoding information about a predicate and the objects it applies to. These representations could be handled by Deep Sets (Zaheer et al., 2017). This would allow to target partial sets of predicates that would not need to characterize all relations between all objects, thus preventing the curse of dimensionality.

Conclusion In this work, we have shown that introducing abstract goals based on relational predicates that are well understood by humans can serve as a pivotal representation between sensorimotor learning of goal-directed capabilities and interaction with a user through language. Here, the role of the caregiver was limited to: 1) helping the agent to experience non-trivial configurations and 2) describing the agent's behavior in a simplified language. In the future, we intend to leverage social interactions to guide agents in the design of their own semantic representations via the use of demonstrations, goal emulations etc.

ACKNOWLEDGMENTS

Anonymized for submission

REFERENCES

Muhannad Alomari, Paul Duckworth, David C Hogg, and Anthony G Cohn. Natural language acquisition and grounding for embodied robotic systems. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

Marcin Andrychowicz, Filip Wolski, Alex Ray, Jonas Schneider, Rachel Fong, Peter Welinder, Bob McGrew, Josh Tobin, Pieter Abbeel, and Wojciech Zaremba. Hindsight Experience Replay. *arXiv preprint arXiv:1707.01495*, 2017.

Dzmitry Bahdanau, Felix Hill, Jan Leike, Edward Hughes, Arian Hosseini, Pushmeet Kohli, and Edward Grefenstette. Learning to understand goal specifications by modelling reward. *arXiv preprint arXiv:1806.01946*, 2018.

Peter W Battaglia, Jessica B Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, et al. Relational inductive biases, deep learning, and graph networks. *arXiv preprint arXiv:1806.01261*, 2018.

Jerome Bruner. Culture, mind, and education. *Contemporary theories of learning*, pp. 159–168, 2009.

Angelo Cangelosi, Giorgio Metta, Gerhard Sagerer, Stefano Nolfi, Chrystopher Nehaniv, Kerstin Fischer, Jun Tani, Tony Belpaeme, Giulio Sandini, Francesco Nori, et al. Integration of action and language knowledge: A roadmap for developmental robotics. *IEEE Transactions on Autonomous Mental Development*, 2(3):167–195, 2010.

Harris Chan, Yuhuai Wu, Jamie Kiros, Sanja Fidler, and Jimmy Ba. Actrce: Augmenting experience via teacher's advice for multi-goal reinforcement learning. *arXiv preprint arXiv:1902.04546*, 2019.

Geoffrey Cideron, Mathieu Seurin, Florian Strub, and Olivier Pietquin. Self-educated language agent with hindsight experience replay for instruction following. *arXiv preprint arXiv:1910.09451*, 2019.

Cédric Colas, Pierre-Yves Oudeyer, Olivier Sigaud, Pierre Fournier, and Mohamed Chetouani. CURIOUS: Intrinsically motivated multi-task, multi-goal reinforcement learning. In *International Conference on Machine Learning (ICML)*, pp. 1331–1340, 2019a.

Cédric Colas, Olivier Sigaud, and Pierre-Yves Oudeyer. A hitchhiker's guide to statistical comparisons of reinforcement learning algorithms. *arXiv preprint arXiv:1904.06979*, 2019b.

Cédric Colas, Tristan Karch, Nicolas Lair, Jean-Michel Dussoux, Clément Moulin-Frier, Peter Ford Dominey, and Pierre-Yves Oudeyer. Language as a cognitive tool to imagine goals in curiosity-driven exploration. *arXiv preprint arXiv:2002.09253*, 2020.

Marc Peter Deisenroth, Carl Edward Rasmussen, and Dieter Fox. Learning to control a low-cost manipulator using data-efficient reinforcement learning. *Robotics: Science and Systems VII*, pp. 57–64, 2011.

Carlos Florensa, David Held, Markus Wulfmeier, Michael Zhang, and Pieter Abbeel. Reverse curriculum generation for reinforcement learning. *arXiv preprint arXiv:1707.05300*, 2017.

Carlos Florensa, David Held, Xinyang Geng, and Pieter Abbeel. Automatic goal generation for reinforcement learning agents. In *International Conference on Machine Learning*, pp. 1514–1523, 2018.

Sébastien Forestier, Yoan Mollard, and Pierre-Yves Oudeyer. Intrinsically motivated goal exploration processes with automatic curriculum learning. *arXiv preprint arXiv:1708.02190*, 2017.

Pierre Fournier, Cédric Colas, Mohamed Chetouani, and Olivier Sigaud. Clic: Curriculum learning and imitation for object control in non-rewarding environments. *IEEE Transactions on Cognitive and Developmental Systems*, 2019.

Justin Fu, Anoop Korattikara, Sergey Levine, and Sergio Guadarrama. From language to goals: Inverse reinforcement learning for vision-based instruction following. *arXiv preprint arXiv:1902.07742*, 2019.

Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. *arXiv preprint arXiv:1801.01290*, 2018.

Steve Harnad. The symbol grounding problem. *Physica D*, 42:335–346, 1990.

Karl Moritz Hermann, Felix Hill, Simon Green, Fumin Wang, Ryan Faulkner, Hubert Soyer, David Szepesvari, Wojciech Marian Czarnecki, Max Jaderberg, Denis Teplyashin, et al. Grounded language learning in a simulated 3d world. *arXiv preprint arXiv:1706.06551*, 2017.

Yiding Jiang, Shixiang Shane Gu, Kevin P Murphy, and Chelsea Finn. Language as an abstraction for hierarchical deep reinforcement learning. In *Advances in Neural Information Processing Systems*, pp. 9414–9426, 2019.

Tristan Karch, Cédric Colas, Laetitia Teodorescu, Clément Moulin-Frier, and Pierre-Yves Oudeyer. Deep sets for generalization in RL. *arXiv preprint arXiv:2003.09443*, 2020.

Johannes Kulick, Marc Toussaint, Tobias Lang, and Manuel Lopes. Active learning for teaching a robot grounded relational symbols. In *Twenty-Third International Joint Conference on Artificial Intelligence*, 2013.

John B. Lanier, Stephen McAleer, and Pierre Baldi. Curiosity-driven multi-criteria hindsight experience replay. *CoRR*, abs/1906.03710, 2019. URL `http://arxiv.org/abs/1906.03710`.

Richard Li, Allan Jabri, Trevor Darrell, and Pulkit Agrawal. Towards practical multi-object manipulation using relational reinforcement learning. *arXiv preprint arXiv:1912.11032*, 2019.

Jelena Luketina, Nantas Nardelli, Gregory Farquhar, Jakob Foerster, Jacob Andreas, Edward Grefenstette, Shimon Whiteson, and Tim Rocktäschel. A survey of reinforcement learning informed by natural language. *arXiv preprint arXiv:1906.03926*, 2019.

Max Lungarella, Giorgio Metta, Rolf Pfeifer, and Giulio Sandini. Developmental robotics: a survey. *Connection Science*, 15(4):151–190, 2003.

Corey Lynch and Pierre Sermanet. Grounding language in play. *arXiv preprint arXiv:2005.07648*, 2020.

Jean M Mandler. On the spatial foundations of the conceptual system and its enrichment. *Cognitive science*, 36(3):421–451, 2012.

Ashvin Nair, Shikhar Bahl, Alexander Khazatsky, Vitchyr Pong, Glen Berseth, and Sergey Levine. Contextual imagined goals for self-supervised robotic learning. *arXiv preprint arXiv:1910.11670*, 2019.

Ashvin V Nair, Vitchyr Pong, Murtaza Dalal, Shikhar Bahl, Steven Lin, and Sergey Levine. Visual reinforcement learning with imagined goals. In *Advances in Neural Information Processing Systems*, pp. 9191–9200, 2018.

Jean Piaget. *The development of thought: Equilibration of cognitive structures.* Viking, 1977. (Trans A. Rosin).

Matthias Plappert, Marcin Andrychowicz, Alex Ray, Bob McGrew, Bowen Baker, Glenn Powell, Jonas Schneider, Josh Tobin, Maciek Chociej, Peter Welinder, et al. Multi-goal reinforcement learning: Challenging robotics environments and request for research. *arXiv preprint arXiv:1802.09464*, 2018.

Vitchyr H Pong, Murtaza Dalal, Steven Lin, Ashvin Nair, Shikhar Bahl, and Sergey Levine. Skew-fit: State-covering self-supervised reinforcement learning. *arXiv preprint arXiv:1903.03698*, 2019.

Rémy Portelas, Cédric Colas, Lilian Weng, Katja Hofmann, and Pierre-Yves Oudeyer. Automatic curriculum learning for deep rl: A short survey. *arXiv preprint arXiv:2003.04664*, 2020.

Sebastien Racaniere, Andrew K Lampinen, Adam Santoro, David P Reichert, Vlad Firoiu, and Timothy P Lillicrap. Automated curricula through setter-solver interactions. *arXiv preprint arXiv:1909.12892*, 2019.

Tom Schaul, Daniel Horgan, Karol Gregor, and David Silver. Universal value function approximators. In *International Conference on Machine Learning*, pp. 1312–1320, 2015.

Robert S Siegler. *Emerging minds: The process of change in children's thinking*. Oxford University Press, 1998.

Olivier Sigaud and Alain Droniou. Towards deep developmental learning. *IEEE Transactions on Cognitive and Developmental Systems*, 8(2):99–114, 2016. doi: 10.1109/TAMD.2015.2496248.

Linda Smith and Michael Gasser. The development of embodied cognition: Six lessons from babies. *Artificial life*, 11(1-2):13–29, 2005.

Kihyuk Sohn, Honglak Lee, and Xinchen Yan. Learning structured output representation using deep conditional generative models. In *Advances in neural information processing systems*, pp. 3483–3491, 2015.

Gerald J. Sussman. A computational model of skill acquisition. Technical report, HIT Technical Report AI TR-297, 1973.

Tadahiro Taniguchi, Emre Ugur, Matej Hoffmann, Lorenzo Jamone, Takayuki Nagai, Benjamin Rosman, Toshihiko Matsuka, Naoto Iwahashi, Erhan Oztop, Justus Piater, et al. Symbol emergence in cognitive developmental systems: a survey. *IEEE Transactions on Cognitive and Developmental Systems*, 11(4):494–516, 2018.

Austin Tate et al. Interacting goals and their use. In *IJCAI*, volume 10, pp. 215–218, 1975.

Stefanie Tellex, Thomas Kollar, Steven Dickerson, Matthew R Walter, Ashis Gopal Banerjee, Seth Teller, and Nicholas Roy. Approaching the symbol grounding problem with probabilistic graphical models. *AI magazine*, 32(4):64–76, 2011.

Emanuel Todorov, Tom Erez, and Yuval Tassa. Mujoco: A physics engine for model-based control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5026–5033. IEEE, 2012.

L. S. Vygotsky. Tool and Symbol in Child Development. In *Mind in Society*, chapter Tool and Symbol in Child Development, pp. 19–30. Harvard University Press, 1978. ISBN 0674576292. doi: 10.2307/j.ctvjf9vz4.6.

Nicholas Waytowich, Sean L Barton, Vernon Lawhern, Ethan Stump, and Garrett Warnell. Grounding natural language commands to starcraft ii game states for narration-guided reinforcement learning. In *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*, volume 11006, pp. 110060S. International Society for Optics and Photonics, 2019.

J. Weng, J. McClelland, A. Pentland, O. Sporns, I. Stockman, M. Sur, and E. Thelen. Autonomous mental development by robots and animals. *Science*, 291(5504):599–600, 2001.

David Wood, Jerome S. Bruner, and Gail Ross. The role of tutoring in problem solving. *Journal of child psychology and psychiatry*, 17(2):89–100, 1976.

Danfei Xu, Suraj Nair, Yuke Zhu, Julian Gao, Animesh Garg, Li Fei-Fei, and Silvio Savarese. Neural task programming: Learning to generalize across hierarchical tasks. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1–8. IEEE, 2018.

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander J Smola. Deep sets. In *Advances in neural information processing systems*, pp. 3391–3401, 2017.

Victor Zhong, Tim Rocktäschel, and Edward Grefenstette. RTFM: Generalising to novel environment dynamics via reading. *arXiv preprint arXiv:1910.08210*, 2019.