Contrastive Goal Representation Learning for Open-Ended Reinforcement Learning

Fabian Wurzberger, Sebastian Gottwald, Daniel Braun

Institute for Neural Information Processing, Ulm University
fabian.wurzberger@uni-ulm.de

Goal-conditioned reinforcement learning lends itself as a natural framework for open-ended reinforcement learning in the absence of external rewards. When agents try to understand their environment by taking exploratory actions, they might choose to recreate some of their experience by setting themselves goals that are neither too simple nor too difficult given the experience so far. Over time the set of possible experiences grows and so does the set of possible goals. A key question in such a setting is, however, how to define goals and how to represent them. In the absence of reward functions, goals are often defined in terms of possible observations, for example an image that reflects the desired final state. However, such observational goals may be at the same time too concrete (we may not care about the background in the image) and too abstract (if all corners look the same, we might want to represent a specific one). This suggests to look at latent representations of sequences of observations and actions that allow both to disambiguate single observations and to abstract away from them.

Recently, contrastive learning has demonstrated significant potential for self-supervised representation learning [2, 4]. In particular, for sequential problems temporal contrastive learning has been proposed, where sequential observations can be viewed as natural augmentations [5]. Moreover, contrastive learning has recently also been employed in the context of goal-conditioned RL where successor feature representations of temporally contiguous states are learnt contrastively [3], but where goals are defined in observation space.

Here, we propose a novel contrastive representation learning approach that focuses specifically on learning latent goals from sequences of observations. In particular, we use recurrent neural networks to encode arbitrary-length sequences $o_{t-\tau} \dots o_t$ into a shared discrete latent representation g, enabling the learning of powerful goal representations capable of encoding both disambiguated positions and abstract regions in navigation-based tasks. We encode hierarchical and temporal structure directly into the latent space through a learnable energy function E which assigns low values to goals (g_a, g_b) if g_a originates from a subset of the observation sequence encoded as g_b. This effectively induces a preorder $\preceq$ on goal space such that $g_a \preceq g_b$ if $E(g_a, g_b) \approx 0$. The energy function provides a natural mechanism for describing set operations such as unions and intersections through optimization, corresponding to infima and suprema of the underlying preorder. The sequential nature of our encoding approach enhances the representation of abstract spatial concepts, moving beyond simple positional information to capture more complex spatial relationships and patterns. Depending on the observation space, encoding specific and abstract goals has a different meaning as shown in Figure 1. Our approach is able to deal with this by incorporating both, specific and abstract goals in the same latent space.

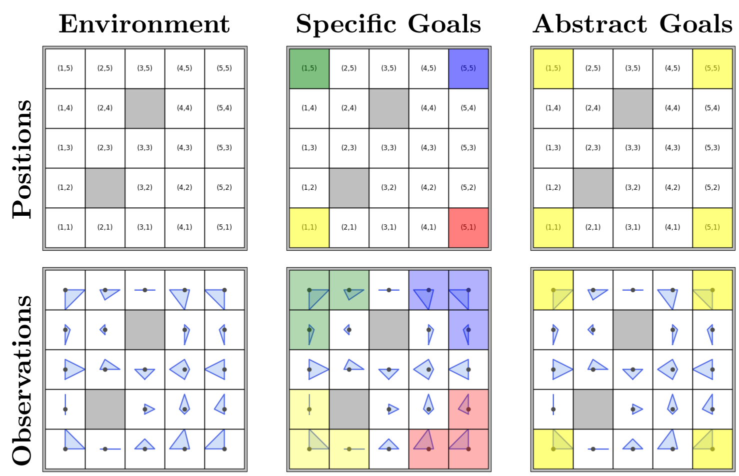

Figure 1: GridWorld environments with different observation types. While positions naturally encode specific goals, abstract goals can not be described as a single position. Similarly, observations are usually ambiguous and thus inherently abstract, however encoding specific goals with a single observation is not possible.

Similar to zero-shot RL [1, 6], we evaluate our goal representation through a two-phase approach where agents are first trained on predefined goal-conditioned tasks using our learned representations, then tested with novel reward functions that are translated into our latent goal space. This evaluation strategy allows us to assess the transferability and generalization capabilities of our learned goal representations across diverse task specifications, demonstrating that the sequential nature of our encoding helps with encoding more abstract spatial concepts. Our approach addresses fundamental challenges in reward-free RL by providing a principled method for learning rich, goal representations that can generalize to novel tasks without additional reward engineering.

References

[1] D. Borsa, A. Barreto, J. Quan, D. J. Mankowitz, H. van Hasselt, R. Munos, D. Silver, and T. Schaul. Universal successor features approximators. In *International Conference on Learning Representations*, 2019.

[2] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton. A simple framework for contrastive learning of visual representations. In *Proceedings of the 37th International Conference on Machine Learning*, 2020.

[3] B. Eysenbach, T. Zhang, S. Levine, and R. R. Salakhutdinov. Contrastive learning as goal-conditioned reinforcement learning. In *Advances in Neural Information Processing Systems*, 2022.

[4] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, E. Buchatskaya, C. Doersch, B. Avila Pires, Z. Guo, M. Gheshlaghi Azar, B. Piot, k. kavukcuoglu, R. Munos, and M. Valko. Bootstrap your own latent - a new approach to self-supervised learning. In *Advances in Neural Information Processing Systems*, 2020.

[5] F. Schneider, X. Xu, M. R. Ernst, Z. Yu, and J. Triesch. Contrastive learning through time. In *SVRHM 2021 Workshop @ NeurIPS*, 2021.

[6] A. Touati and Y. Ollivier. Learning one representation to optimize all rewards. In *Advances in Neural Information Processing Systems*, 2021.