Deep Symbolic Optimization for Scientific Discovery

Mikel Landajuela, Claudio Santiago, T. Nathan Mundhenk, Chak Shing Lee, Jacob Pettit, Jiachen Yang, Ruben Glatt, Ignacio Aravena, Felipe Leno Da Silva, Andre Goncalves, Alexander Ladd, Daniel Faissol, Brenden Petersen

Lawrence Livermore National Laboratory, Livermore, CA, USA

January 14, 2023

In this work, we present Deep Symbolic Optimization (DSO), a computational framework for scientific discovery [3, 1]. DSO treats the discovery problem as a sequential decision-making task and learns a probabilistic model over the space of candidate designs (e.g., mathematical expressions, control policies, amino acid sequences). The probabilistic model, which is parameterized by a generative neural network, is then optimized to allocate the search effort in the most promising regions of the design space. The optimization process combines the use of gradient-based techniques (inspired by the reinforcement learning literature) with evolutionary search methods and/or local search techniques. DSO allows for the incorporation of *in-situ* constraints, advanced policy optimization techniques, pre-training the generative model on related discovery problems, and the use of powerful deep learning frameworks for online learning. DSO can also be provided with additional knowledge in the form of priors or additional representations from Large Language Models that provide better traction and guidance in the search process. The versatility of DSO has been demonstrated through its successful application to, among other problems, the discovery of mathematical models of data [1] and the discovery of interpretable policies in stochastic control problems [2]. In the case of mathematical discovery problems, DSO has been show to achieve state-of-the-art performance in terms of accuracy and interpretability in numerous benchmark problems and competitions. In the case of control problems, DSO has been shown to discover policies that are competitive with state-of-the-art neural network policies but with the added benefit of being interpretable and explainable [2]. Current efforts are focused on further development of DSO for discovering antibody sequences that bind to specific targets based on large-scale molecular simulations, data-driven predictive models, and curated experimental datasets. In this talk, we will describe in detail the DSO framework and its applications, as well as discuss future directions for this promising approach to scientific discovery.

[1] M. Landajuela, C. Lee, J. Yang, R. Glatt, C. P. Santiago, I. Aravena, T. N. Mundhenk, G. Mulcahy, and B. K. Petersen. A unified framework for deep symbolic regression. In *Advances in Neural Information Processing Systems*, 2022.

[2] M. Landajuela, B. K. Petersen, S. Kim, C. P. Santiago, R. Glatt, N. Mundhenk, J. F. Pettit, and D. Faissol. Discovering symbolic policies with deep reinforcement learning. In M. Meila and T. Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 5979–5989. PMLR, 18–24 Jul 2021.

[3] B. K. Petersen, M. Landajuela, T. N. Mundhenk, C. P. Santiago, S. Kim, and J. T. Kim. Deep symbolic regression: Recovering mathematical expressions from data via risk-seeking policy gradients. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021.