Multi-objective Materials Discovery using Weighted Preference Optimization

Andrew H. Salij[1], Cristina Garcia Cardona[1], Ivana Matanovic[1], Megan C. Davis[1], R. Seaton Ullberg[1], Marc J. Cawkwell[1], Christopher Snyder[1], and Wilton Jr. M. Kort-Kamp[1]

[1]Los Alamos National Laboratory, P.O. Box 1663, Los Alamos, NM, United States of America

asalij@lanl.gov

Generative artificial intelligence, particularly in domain specific models such as chemical language models, has demonstrated great promise for the initial stage of candidate generation in materials discovery [1]. To identify desirable new materials, one must generally balance a variety of characteristics, but common approaches to discovery problems using generative AI optimize for a particular metric or reward of interest [2, 3]. Such a singular focus threatens to degrade material performance across unconstrained metrics as the model optimizes for one goal at the expense of other qualities.

Recent advances in language model development, particularly in post-training, have highlighted the benefits for model alignment of directly training on preference data in Direct Preference Optimization (DPO) [4]. Such a direct approach avoids the need for an explicit reward model, providing a robust means of improving language models to a preferred goal. While DPO has become heavily used in natural language processing, it remains underexplored for chemical language modelling. The fact that molecules can be automatically characterized presents opportunities for constructing preference data that do not exist for natural language data, making DPO of particular interest for materials discovery.

We introduce a general method for automatically constructing preference data for molecular discovery that balances across a variety of objectives, providing an autonomous approach for fine-tuning generative models for targeted generation when coupled with preference optimization training. This Iterative Weighted Schema Weighted Preference Optimization (IWSDPO) extends prior approaches to optimizing across desired characteristics and presents a means of creating material candidates that can balance even competing characteristics.

We demonstrate the effectiveness of training in an iterative loop, making the reinforcement learning training into more of an online methodology than traditional DPO. Additionally, we highlight how our novel introduction of weighted schemas produces an implicit multi-objective reward without need for a single reward function. Such multi-objective training enables improvement in model molecular generation across a variety of metrics and in multiple regions of chemical space.

This talk will provide a high-level mathematical intuition as to why such direct training using a weighted schema works as well as demonstrate cases for materials discovery such as carbon sinks.

References

[1] Francesca Grisoni, *Current Opinion in Structural Biology* 79 (2023) 102527.
[2] Gavin Ye, *Journal of Computer-Aided Molecular Design* 38 (2024) 20.
[3] Pablo V. Pardo, O. Toledano, Guillermo Marcos-Ayuso, D. Quesada, N. E. Campillo, *Journal of Chemical Information and Modeling* 66 (2026), 910-922.
[4] Rafael Rafailov, Archit Sharma, Eric Mitchell, Chris D. Manning, Stefano Ermon, Chelsea Finn, *Advances in Neural Information Processing Systems*, 36 (2023), 53728-53741.

Figures

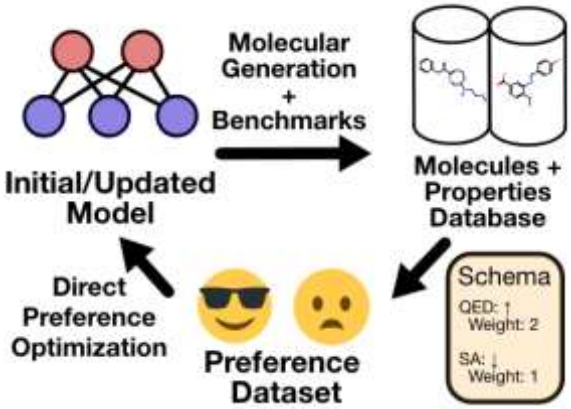

Figure 1. General schema for Iterative Weighted Schema Direct Preference Optimization. Smiley Face: (CC-BY 4.0) Twitter. Unhappy Face: (CC-BY 4.0) Twitter.