Learning & Adaptive Systems

ETH zürich

De Novo Molecular Design via Diffusion Bandit Optimization

Project Proposal for Master Thesis

Motivation

Designing molecules with targeted properties is crucial for applications ranging from drug-design to designing sustainable chemical processes [Bilodeau et al., 2022]. Recently, generative models such as diffusion or flow matching models have succeeded in producing molecules with similarity to existing chemical datasets [Hoogeboom et al., 2022, Runcie and Mey, 2023]. While diffusion models are promising to sample complex (high-dimensional or combinatorial) spaces, they do not naturally lead to designs optimizing a specific property via online feedback. Meanwhile, Bayesian optimization techniques lead to good property maximizers but do not scale easily to complex domains. Hence the question: can we combine the ability of generative models to produce promising molecules while achieving property improvements over the sequence of generated molecules according to a notion of optimality? We have recently started to answer this question by leveraging diffusion models similarly to [Yuan et al., 2024, Uehara et al., 2024] and designed algorithms for Bayesian optimization via diffusion models. Now we aim to specialize these ideas into a practical method and test it on a real-world de novo molecular design problem.

Scope of the Project

Starting from a principled algorithm for generative Bayesian optimization, we wish to build a more practical method and adapt it for real-world molecular design problems.

Ideal Candidate

The project is focused on applying and adapting the derived methodology to molecular design, without emphasis on theoretical derivations. A good candidate will have a mix of the following.

- Understanding of Bayesian optimization and/or diffusion models.
- Coding experience with Bayesian optimization and/or diffusion (generative) models.

Notice that a background in molecular design or chemistry is not required, however a willingness to learn working with molecular datasets and equivariant diffusion models is crucial.

Contact

If you are interested, please contact Riccardo De Santi (rdesanti@ethz.ch) and Luca Schaufelberger (schaluca@ethz.ch) with a short description of your motivation and a CV.

This project is a collaboration between the Learning & Adaptive Systems Group (Prof. Andreas Krause) and the Digital Chemistry Laboratory (Prof. Kjell Jorner).

References

[Bilodeau et al., 2022] Bilodeau, C., Jin, W., Jaakkola, T., Barzilay, R., and Jensen, K. F. (2022). Generative models for molecular discovery: Recent advances and challenges. *WIREs Computational Molecular Science*, 12(5):e1608. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1608.

[Hoogeboom et al., 2022] Hoogeboom, E., Satorras, V. G., Vignac, C., and Welling, M. (2022). Equivariant Diffusion for Molecule Generation in 3D. In *Proceedings of the 39th International Conference on Machine Learning*, pages 8867–8887. PMLR. ISSN: 2640-3498.

[Runcie and Mey, 2023] Runcie, N. T. and Mey, A. S. (2023). SILVR: Guided Diffusion for Molecule Generation. *Journal of Chemical Information and Modeling*, 63(19):5996–6005. Publisher: American Chemical Society.

[Uehara et al., 2024] Uehara, M., Zhao, Y., Black, K., Hajiramezanali, E., Scalia, G., Diamant, N. L., Tseng, A. M., Levine, S., and Biancalani, T. (2024). Feedback efficient online fine-tuning of diffusion models. *arXiv preprint arXiv:2402.16359*.

[Yuan et al., 2024] Yuan, H., Huang, K., Ni, C., Chen, M., and Wang, M. (2024). Reward-directed conditional diffusion: Provable distribution estimation and reward improvement. *Advances in Neural Information Processing Systems*, 36.