Research statement

My researching interests are mainly (1) **diffusion model for combinatorial optimization problems** and (2) **neural architecture search (NAS)**, especially zero-shot NAS.

Diffusion model for combinatorial optimization problems

This topic attracts my attention because of this paper [6]. In this work, Sun and Yang leverage graph-based diffusion model to solve travelling salesman problem and maximal independent set. I had cooperated with my research supervisor, Tsai, and one of his student, Wu, to improve the model. We tried to extend this model to vehicle routing problem and Dial-a-Ride problem, and even neural architecture search. Wu proposed the loss function should include not only supervised loss but also a local loss which takes the difference between the ground truth routes and the predicted routes into account. The result shows that, compare to the other SOTA methods, the proposed method can outperform others "when they are trained and tested on data of the same scale." In other words, the scalability of the proposed method is weaker than others. The cause may hide behind the local loss. Since the local loss is discrete, it cannot directly optimize the model. And the local loss influences the searching process of optimizer, which turns out that the model overfit on the training scale. This problem probably can be solved by utilizing the reinforcement learning or other techniques. I think the it is also worth developed because of its versatility. For example, it may solve other combinatorial problems or even neural architecture search. Moreover, the recent papers, [7] (leverage diffusion model to help heuristic algorithm) and [8] (leverage diffusion model to help the policy-regularization in offline reinforcement learning), also show that diffusion model is worth studied.

Neural architecture search

The classic approach to perform NAS is use reinforcement learning [1], or meta-heuristic algorithm [2]. The training cost is still high because a candidate architecture must be trained to evaluate during the searching process. This shortcoming leads to the idea of few-shot NAS [3] and zero-shot NAS [4]. Zero-shot NAS relies on training-free metrics, or called zero-cost proxy. Lots of the metrics are proposed in recent years. However, most of the publications use only one metrics to evaluate a neural networks. Therefore, in my publication, I tried to combine different metrics and the result shows combination of different metrics does improve the search performance. During the research, I also found that different combinations of metrics have different improvement to the performance. The architecture can be seen as a box. And the metrics are the different view of the architecture. The more different views we have, the more information we have. Note that the categories of zero-cost proxies exist [5], so I think the metrics that evaluate the architecture should not be in the same category. Based on this hypothesis, I choose three metrics that evaluate the generalization, expressivity and trainability of the architecture. That's what I have done in my publication. However, a fundamental question raise in my mind, are the metrics really evaluate the generalization, expressivity and trainability of an architecture? Can these metrics be proxies of these properties? One more question, can we first clustering these metrics automatically when utilizing these metrics on different dataset of architectures and then choose the metrics to maximize the performance? I think these questions are worth studied and developed, since it leads to a deeper understanding of how neural networks behave and what makes them better.

References

[1] B. Zoph and Q. Le, "Neural architecture search with reinforcement learning," in *Proceedings of the International Conference on Learning Representations*, 2017.

[2] M.-T. Wu, H.-I. Lin, and C.-W. Tsai, "A training-free neural architecture search algorithm based on search economics," *Evolutionary Computation*, vol. 28, no. 2, pp. 445–459, 2024.

[3] H. Liu, K. Simonyan, and Y. Yang, "DARTS: Differentiable architecture search," in *Proceedings of the International Conference on Learning Representations*, 2019.

[4] J. Mellor, J. Turner, A. Storkey, and E. J. Crowley, "Neural architecture search without training," in *Proceedings of the International Conference on Machine Learning*, vol. 139. PMLR, 2021, pp. 7588–7598.

[5] "A deeper look at zero-cost proxies for lightweight NAS," in *Blog Track at ICLR 2022*, 2022. [Online]. Available: https://openreview.net/forum?id=10XI9NWz2nm

[6] Z. Sun and Y. Yang, "DIFUSCO: Graph-based diffusion solvers for combinatorial optimization," in *Proceedings of the International Conference on Neural Information Processing Systems*, 2023.

[7] L. Xin, W. Song, Z. Cao, and J. Zhang, "NeuroLKH: Combining deep learning model with lin-kernighan-helsgaun heuristic for solving the traveling salesman problem," in *Proceedings of the Neural Information Processing Systems*, 2021.

[8] Z. Wang, J. J. Hunt, and M. Zhou, "Diffusion policies as an expressive policy class for offline reinforcement learning," in *Proceedings of The Eleventh International Conference on Learning Representations*, 2023.