Preference-Driven Multi-Objective Combinatorial Optimization with Conditional Computation

Anonymous Author(s)
Affiliation
Address
`email`

Abstract

Recent deep reinforcement learning methods have achieved remarkable success in solving multi-objective combinatorial optimization problems (MOCOPs) by decomposing them into multiple subproblems, each associated with a specific weight vector. However, these methods typically treat all subproblems equally and solve them using a single model, hindering the effective exploration of the solution space and thus leading to suboptimal performance. To overcome the limitation, we propose POCCO, a novel plug-and-play framework that enables adaptive selection of model structures for subproblems, which are subsequently optimized based on preference signals rather than explicit reward values. Specifically, POCCO integrates a conditional computation block into the decoder, where a sparse gating network dynamically routes each subproblem through either a subset of feed-forward (FF) experts or a parameter-free identity (ID) expert. This enables context-aware selection of computation paths, effectively scaling model capacity and enhancing representation learning. Moreover, POCCO replaces raw scalarized rewards with pairwise preference learning: for each subproblem, the policy samples two trajectories, identifies the preferred one, and optimizes a Bradley–Terry likelihood based on their average log-likelihoods. This comparative feedback guides learning toward more preferred solutions, promoting efficient exploration and faster convergence. We integrate POCCO into two state-of-the-art neural MOCOP solvers—CNH and WE-CA—yielding POCCO-C and POCCO-W, respectively. As shown in Table 1, POCCO-W consistently outperforms WE-CA across all benchmarks, setting a new state-of-the-art among neural MOCOP methods. Similarly, POCCO-C surpasses CNH in every case, demonstrating its clear advantage.

Table 1: Performance on Bi-TSP, Bi-CVRP, Bi-KP, and Tri-TSP Instances

Task	Method	Small			Medium			Large		
		HV	Gap	Time	HV	Gap	Time	HV	Gap	Time
Bi-TSP	CNH	0.6270	0.00%	13s	0.6387	0.48%	16s	0.7019	0.83%	33s
	POCCO-C	0.6275	-0.08%	14s	0.6409	0.14%	20s	0.7047	0.44%	42s
	WE-CA	0.6270	0.00%	6s	0.6392	0.41%	9s	0.7034	0.62%	18s
	POCCO-W	0.6275	-0.08%	7s	0.6411	0.11%	14s	0.7055	0.32%	36s
Bi-CVRP	CNH	0.4287	0.33%	11s	0.4087	0.51%	15s	0.4065	0.59%	25s
	POCCO-C	0.4294	0.16%	16s	0.4101	0.17%	25s	0.4079	0.24%	53s
	WE-CA	0.4290	0.26%	7s	0.4089	0.46%	10s	0.4068	0.51%	21s
	POCCO-W	0.4294	0.16%	8s	0.4102	0.15%	17s	0.4084	0.12%	46s
Bi-KP	CNH	0.3556	0.17%	16s	0.4527	0.15%	23s	0.3598	0.14%	55s
	POCCO-C	0.3560	0.06%	20s	**0.4535**	-0.02%	36s	**0.3603**	0.00%	1.4m
	WE-CA	0.3558	0.11%	8s	0.4531	0.07%	16s	0.3602	0.03%	50s
	POCCO-W	**0.3562**	0.00%	11s	0.4534	0.00%	26s	**0.3603**	0.00%	1.3m
Tri-TSP	CNH	0.4698	0.30%	10s	0.4358	1.78%	14s	0.4931	2.32%	25s
	POCCO-C	0.4704	0.17%	18s	0.4393	0.99%	17s	0.4985	1.25%	28s
	WE-CA	0.4707	0.11%	5s	0.4389	1.08%	8s	0.4975	1.45%	17s
	POCCO-W	0.4710	0.04%	6s	0.4397	0.90%	13s	0.4985	1.25%	23s

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.