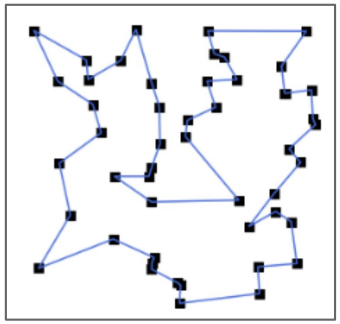
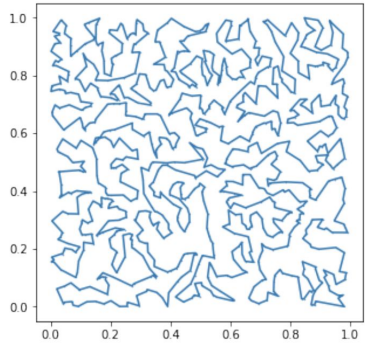
Combinatorial Optimization by Graph Pointer Networks and Hierarchical Reinforcement Learning

Qiang Ma, Suwen Ge, Danyang He, Darshan Thaker, Iddo Drori

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

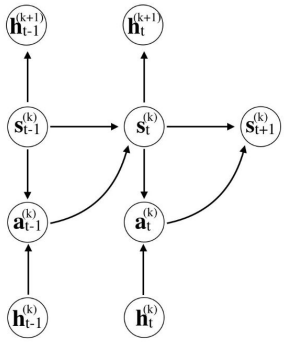
Data, Models, and Code: https://github.com/qiang-ma/graph-pointer-network



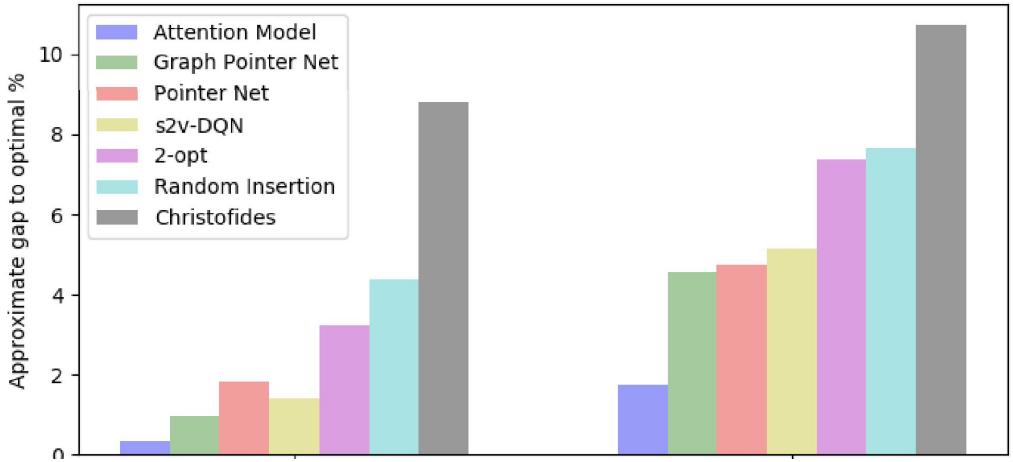


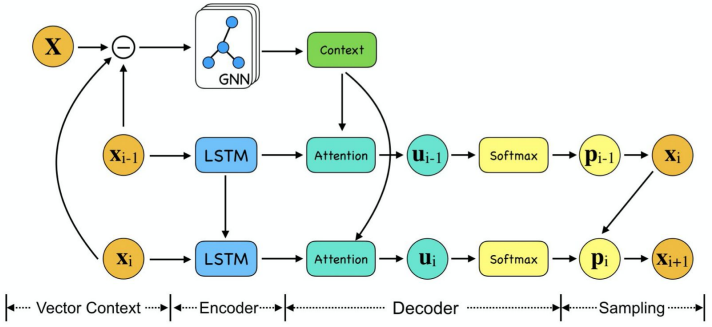
Train on small-scale problems (TSP 50) **generalize well to large-scale problems (TSP 1000)**

Hierarchical Reinforcement Learning



Graphical models for hierarchical RL framework. The next action is conditioned both on the current state and the latent variable from the lower layer. It also provides the latent variable for the next higher layer. we set lower layer reward functions to simply bias solutions to be in the feasible set of the constrained optimization problem, and set higher layer reward functions to be the original optimization objective.





Architecture for GPN. The current city coordinate is encoded by LSTM. All city coordinates are encoded by a GNN. The encoded features are passed to the attention decoder, which outputs a probability distribution over the next candidate cities.

Method	TSP 250		TSP 500		TSP 750		TSP 1000	
	Tour Len.	Time	Tour Len.	Time	Tour Len.	Time	Tour Len.	Time
LKH	11.893	9792s	16.542	23070s	20.129	36840s	23.130	50680s
Concorde	11.89	1894s	16.55	13902s	20.10	32993s	23.11	47804s
Nearest Neighbor	14.928	25s	20.791	60s	25.219	115s	28.973	136s
2-opt	13.253	303s	18.600	1363s	22.668	3296s	26.111	6153s
Farthest Insertion	13.026	33s	18.288	160s	22.342	454s	25.741	945s
OR-Tools (Savings)	12.652	5000s	17.653	5000s	22.933	5000s	28.332	5000s
OR-Tools (Christofides)	12.289	5000s	17.449	5000s	22.395	5000s	26.477	5000s
s2v-DQN	13.079	476s	18.428	1508s	22.550	3182s	26.046	5600s
Pointer Net	14.249	29s	21.409	280s	27.382	782s	32.714	3133s
Attention Model	14.032	2s	24.789	14s	28.281	42s	34.055	136s
GPN (ours)	13.679	32s	19.605	111s	24.337	232s	28.471	393s
GPN+2opt (ours)	12.942	214s	18.358	974s	22.541	2278s	26.129	4410s