Learning General Optimal Policies with Graph Neural Networks: Expressive Power, Transparency, and Limits

Simon Ståhlberg
Linköping University, Sweden

Blai Bonet
Universitat Pompeu Fabra, Spain

Hector Geffner
ICREA, Spain
Universitat Pompeu Fabra, Spain
Linköping University, Sweden

Key points

- General optimal policies:
  - Trained in a supervised manner
  - Optimal value functions are targets
  - Training set consists of small instances
- Graph neural networks:
  - Messages are passed along ground atoms instead of edges
  - Models are defined from the predicates of the planning domain
  - Planning states are used as-is as input
- Expressive power:
  - Graph neural networks are bounded by C_2
  - C_2 is a fragment of first-order logic restricted to 2 variables
  - We can only learn a model that generalize properly if the optimal value function can be expressed in C_2
  - We show that the theoretical results and the experimental results align

Learn **general optimal policies** for classical planning that gets close to **100% coverage** with a simple extension to **graph neural networks**

Input: A a set of ground atoms s (state and goal atoms) over a set of objects $\mathcal{O}$
Output: A scalar value v

$f_0(o) \sim \mathbf{0}^{k/2}\mathcal{N}(0,1)^{k/2}$ for each object $o \in \mathcal{O}$;
for $i \in \{0, \dots, L-1\}$ **do**
  for $q := p(o_1, \dots, o_m) \in s$ **do**
    $m_{q,o} := [\mathbf{MLP}_p(f_i(o_1), \dots, f_i(o_m))]_j$;
  for $o \in \mathcal{O}$ **do**
    $f_{i+1}(o) := \mathbf{MLP}_U(f_i(o), \mathbf{agg}(\{\!\{m_{q,o} | q \in s\}\!\}))$;
$v = \mathbf{MLP}_2(\Sigma_{o \in O}\mathbf{MLP}_1(f_L(o)))$

Number of objects in the problems in the training, validation and test datasets.

Domain	Train	Validation	Test
Blocks-clear	[2, 9]	[10, 11]	[12, 17]
Blocks-on	[2, 9]	[10, 11]	[12, 17]
Gripper	[10, 18]	[20, 22]	[24, 48]
Logistics	[17, 24]	[31, 31]	[31, 39]
Miconic	[5, 26]	[29, 35]	[38, 92]
Parking-behind	[21, 27]	[30, 30]	[30, 36]
Parking-curb	[21, 27]	[30, 30]	[30, 36]
Rovers	[15, 52]	[53, 62]	[67, 116]
Satellite	[14, 41]	[47, 59]	[50, 103]
Transport	[14, 39]	[38, 43]	[41, 77]
Visitall	[27, 102]	[102, 146]	[171, 326]

General architecture that outputs a scalar value v for a given state s.

Domain (#)	L	Opt.	Sub.
Blocks-clear (11)	82	11	0
Blocks-on (11)	150	11	0
Gripper (39)	117	39	0
Logistics (8)	48	8	0
Miconic (95)	378	95	0
Parking-behind (32)	77	32	0
Parking-curb (32)	101	32	0
Rovers (26)	111	20	6
Satellite (20)	97	20	0
Transport (20)	208	20	0
Visitall (12)	93	12	0
Total (306)	1,462	300 (98%)	6 (2%)

Number of problems in test set solved optimally, suboptimally, or not solved at all. Total number of problems (#) shown in parenthesis. L is the sum of all optimal plan lengths.