Learning Domain-Independent Policies for Open List Selection

André Biedenkapp, David Speck, Silvan Sievers, Frank Hutter, Marius Lindauer, Jendrik Seipp
Motivation

What state should I expand next?

Planner

\[ h_1 \]

\[ h_2 \]

\[ h_3 \]

\[ h_4 \]
Satisficing planning with multiple heuristics

• Search for a good plan
• Inadmissible heuristics are difficult to combine
• States evaluated with each heuristic
• One separate open list for each heuristic
Is Domain Dependent
Good enough?

Training Domain

barman

barman

Generalization Domain

blockworld

childsnack

driverlog

floortile

visitall

1
Combining Instance and Dynamic Features

select open list $O_{k}^{t+1}$

DAC-policy $\tilde{\pi}$

planner $A$

reward $\tilde{r}_{t}^{i}$

control of $O \in \tilde{\Theta}$

DAC-state $\tilde{s}_{t}^{i}$

instance $i$

preprocessing

feature generator
• Using Only Dynamic Features (no-F)
• Concatenating Instance and Dynamic Features (raw-F)
• Learning Separate Representations (embed)
• Decoupling Instance and Dynamic Features (dc)
<table>
<thead>
<tr>
<th>Training Method</th>
<th>no-F</th>
<th>1</th>
<th>1</th>
<th>0.57</th>
<th>0.99</th>
<th>1.1</th>
<th>0.86</th>
</tr>
</thead>
<tbody>
<tr>
<td>raw-F</td>
<td>0.89</td>
<td>0.97</td>
<td>0.66</td>
<td>0.82</td>
<td>1</td>
<td>0.85</td>
<td></td>
</tr>
<tr>
<td>embed</td>
<td>0.86</td>
<td>0.93</td>
<td>0.75</td>
<td>0.84</td>
<td>1</td>
<td>0.85</td>
<td></td>
</tr>
<tr>
<td>dc</td>
<td>0.99</td>
<td>0.94</td>
<td>0.7</td>
<td>0.96</td>
<td>0.94</td>
<td>0.85</td>
<td></td>
</tr>
</tbody>
</table>

**Generalization Domain**

- barman
- blocksworld
- childsnack
- driverlog
- floortile
- visitall

Color bar:

- 1.00
- 0.75
- 0.50
- 0.00
Looking forward to meeting you at the poster!
Goal: Select configuration for the problem at hand and adapt while planning

Planner

add new states to

Open List for $h_1$ -> state $s$
Open List for $h_2$ -> state $t$
Open List for $h_3$ -> state $u$
Open List for $h_4$ -> state $v$

RL Agent

Instance
Configuration -> Planner -> Performance
Goal: Find single best configuration

Configuration -> Planner -> Performance
Motivation - AS

Goal: Select best configuration for the problem at hand

Configuration + Instance Features

-> Planner -> Performance
Motivation - DAC

Open List for $h_1$ -> state $m$
Open List for $h_2$ -> state $v$
Open List for $h_3$ -> state $x$
Open List for $h_4$ -> state $o$

Planner selects from Instance 1 and evaluates state $o$.
Motivation - DAC

Open List for $h_1$ -> state $l$
Open List for $h_2$ -> state $q$
Open List for $h_3$ -> state $p$
Open List for $h_4$ -> state $v$

select from Planner

evaluate

Instance 2

RL Agent

Biedenkapp, Speck, Sievers, Hutter, Lindauer, Seipp
1. We generalize previous DAC approaches to learn domain-independent open list selection policies

2. We present novel ways to learn from instance specific features jointly with dynamic features

3. Our learned policies reduce the required number of node expansions on several domains

4. We use DAC as a tool to gain insights on why LAMA’s policy has such strong performance