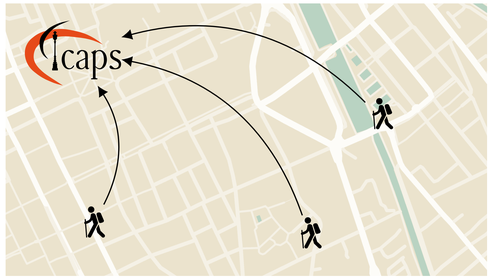
Motivation



Frequently, we have to **solve** tasks from the **same state space** with the **same goal**.

⇒ **Once** learn a good heuristic, speed up future searches.

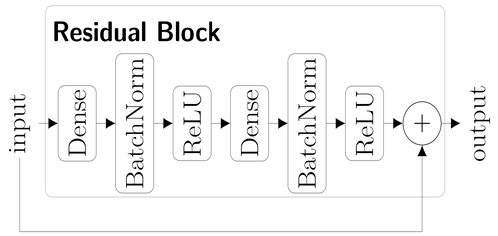
FDR Task

An FDR planning task is a tuple $\Pi = (V, A, I, G)$ with:

- V: a set of multi-valued variables
- A: a set of actions
- I: an initial state
- G: a partial variable assignment

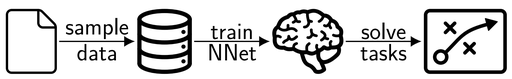
States assign every variable a value.
A state s can be represented as vector:
$[s(v) = d \textbf{ for } v \in V \textbf{ for } d \in dom(V)]$

Residual Network



We use one residual block and two dense layers. Each dense layer has 250 neurons.

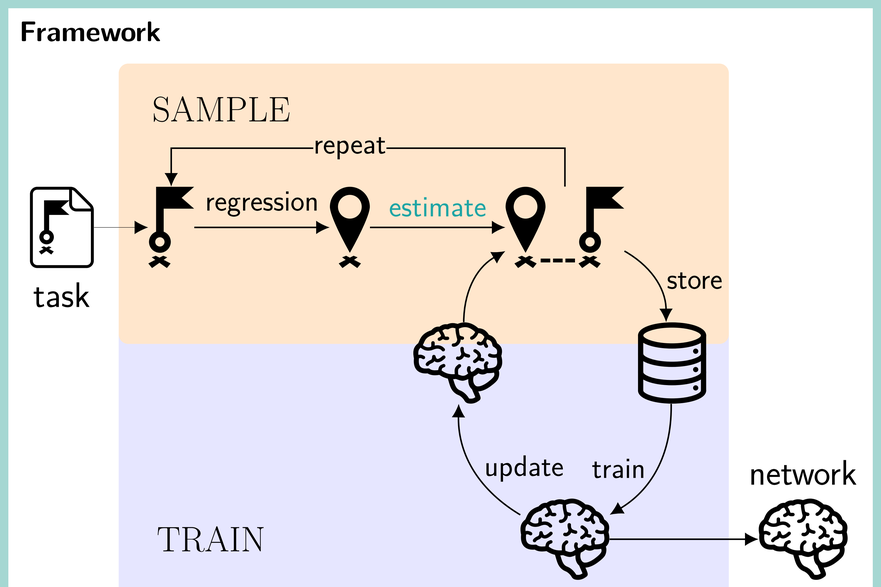
Our Previous Work (SL)



1. Sample for 400 hours data via progression and solve with GBFS(FF).
2. Use supervised learning to train NNet.

Reinforcement Learning can be superior to Supervised Learning **for learning heuristics**.

Framework



Estimate: Approx. Value Iteration (AVI)

$$estimate(\text{📍}) = \min_{x \in \text{neighbors}} (\text{🧠}(x) + distance(\text{📍}, x))$$

Estimate: Sampling Search (SaSe)

- Use the network (🧠) as heuristic[1] to find a **plan**.
- $estimate(\text{📍}) = |\textbf{plan}|$
- Increase the number of regression steps over time.

[1]GBFS with 10s timeout

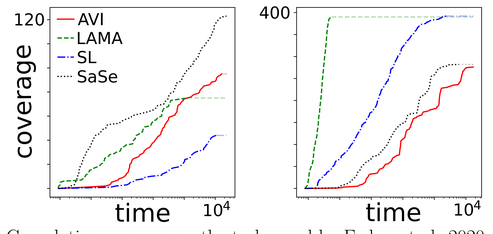
Results

Domain	AVI	SaSe	SL	Lama
blocks	0.0	0.0	**98.0**	96.8
depots	17.7	39.7	64.3	**98.7**
grid	51.0	86.0	74.0	**97.0**
npuzzle	1.0	1.5	0.0	**97.8**
pipesworld-nt	29.8	50.4	92.8	**97.2**
rovers	25.8	35.8	12.5	**98.0**
scanalyzer	83.3	33.3	77.7	**97.7**
storage	47.5	**71.5**	22.0	37.5
transport	69.0	70.5	**98.0**	97.5
visitall	13.0	30.7	0.7	**95.0**
Average	33.8	41.9	54.0	**91.3**

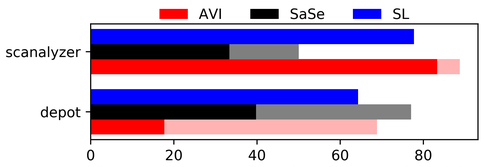
Coverage (in %) on the tasks used by Ferber et al. 2020.

Domain	AVI	SaSe	Lama
depots	15.1	6.9	**80.6**
grid	0.0	0.0	**90.0**
npuzzle	0.0	0.0	**84.0**
pipesworld-nt	1.4	25.1	**68.7**
rovers	0.1	0.8	**97.7**
scanalyzer	34.0	3.3	**98.7**
storage	18.8	**26.5**	11.0
visitall	0.0	36.0	**98.0**
Average	8.7	12.3	**78.6**

Coverage (in %) on the tasks too hard for training data generation for Ferber et al. 2020.



Cumulative coverage on the tasks used by Ferber et al. 2020 for the storage and transport domains.



Coverage (in %) improvements after identifying good models for the tasks by Ferber et al. 2020.

Conclusions & Future Work

- We have no single best NN technique.
- Identifying successful models leads to significant improvements.
- Future Work: Predicting expansions and incorporating unsolved samples of **SaSe**.

Reinforcement Learning for Planning Heuristics

Patrick Ferber, Malte Helmert, and Jörg Hoffmann
University of Basel, Switzerland, and Saarland University, Germany