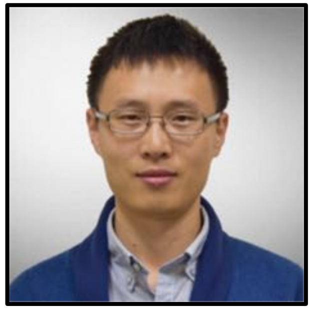
Tencent AI Lab

NeurIPS

Synthesized Policies for Transfer and Adaptation across Tasks and Environments



Hexiang Hu*, Liyu Chen*, Boqing Gong, Fei Sha

NETFLIX

USC Viterbi School of Engineering

Transfer Learning in RL

mop the floor

wash dishes

cooking

In this work we decompose environments and tasks, and consider three progressively more difficult transfer settings

A good household robot needs to complete multiple tasks

USC Viterbi School of Engineering

Image Credits: retrieved from: http://twomaverix.com/?p=14570, https://www.stufftoblowyourmind.com/blogs/robots-will-cook-you-dinner.htm, http://www.mehmetari.com/blog/time-da-yayinlanan-yeni-nesil-robot-goruntuleri/

Transfer Learning in RL

A's home

B's home

C's home

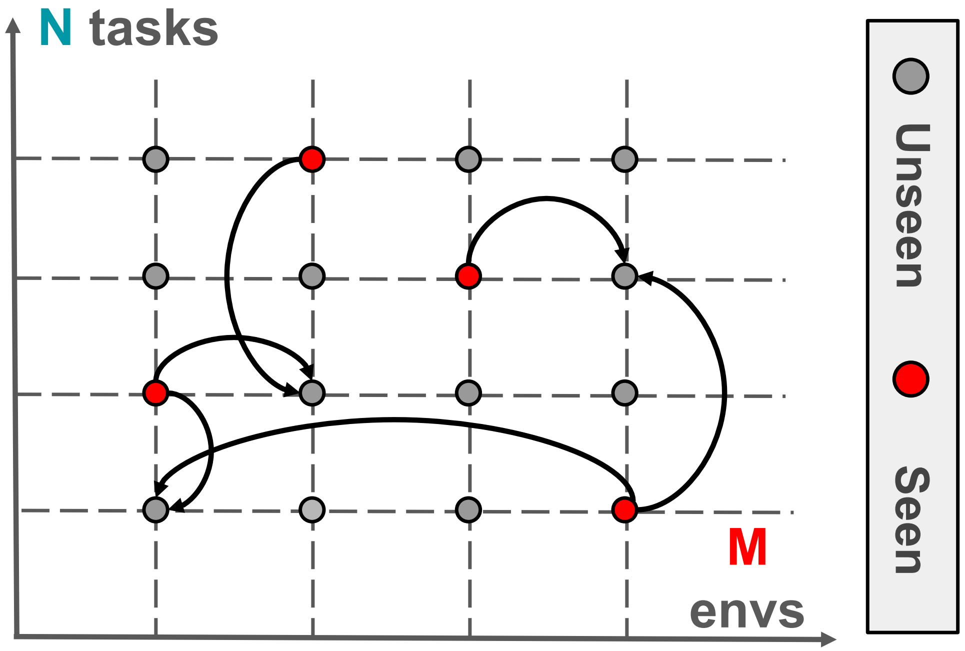
In this work we decompose environments and tasks, and consider three progressively more difficult transfer settings

A good household robot needs to complete multiple tasks in multiple environments

Image Credits: retrieved from: http://v1.escapistmagazine.com/news/view/100397-Humanoid-Dishbot-Will-Force-You-to-Do-Chores, https://www.stufftoblowyourmind.com/blogs/robots-will-cook-you-dinner.htm, https://spielraum.xing.com/2015/03/kollege-roboter-ist-schon-da/

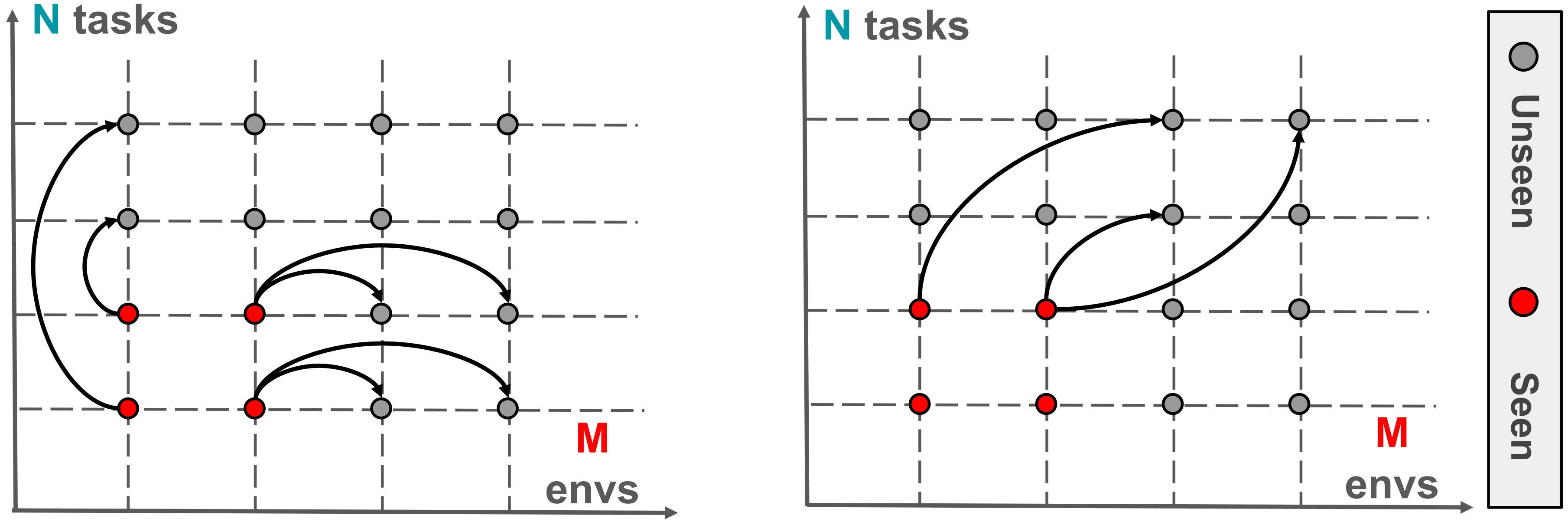
USC Viterbi School of Engineering

Transfer Settings I



- Transfer to a new (env, task) pair, with seen environment and seen task

Transfer Settings 2 & 3

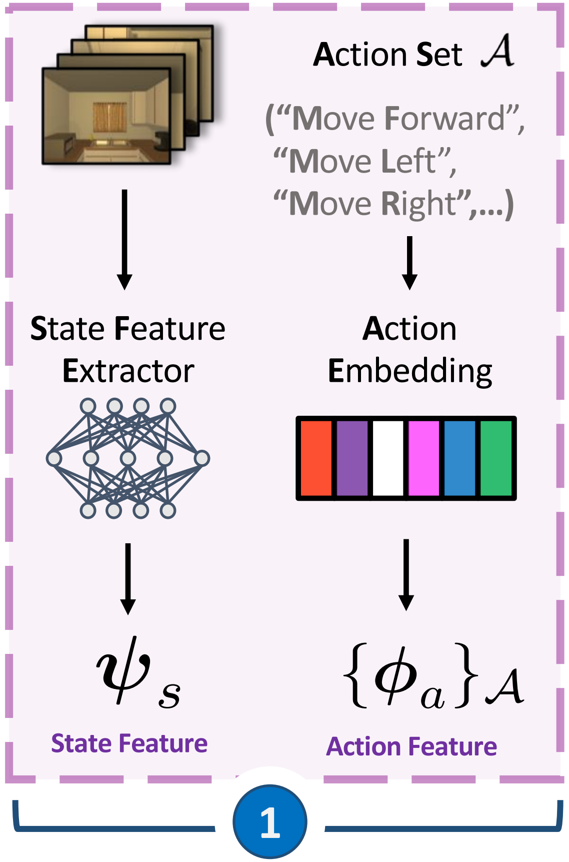


- Transfer to a seen environment and unseen task, or unseen environment and seen task, or unseen environment and unseen task

USC Viterbi
School of Engineering

Our Method

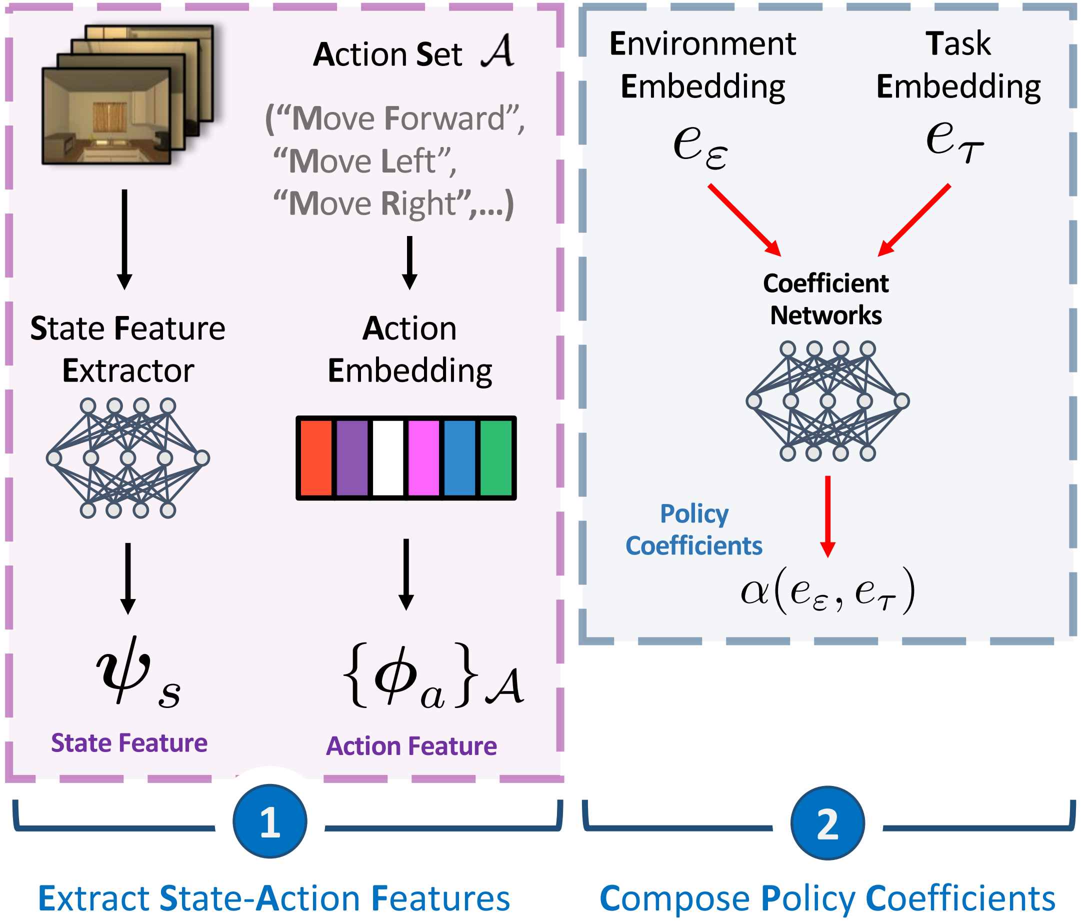
State, action features and policy basis are learned across all seen (env, task) comb.



Extract State-Action Features

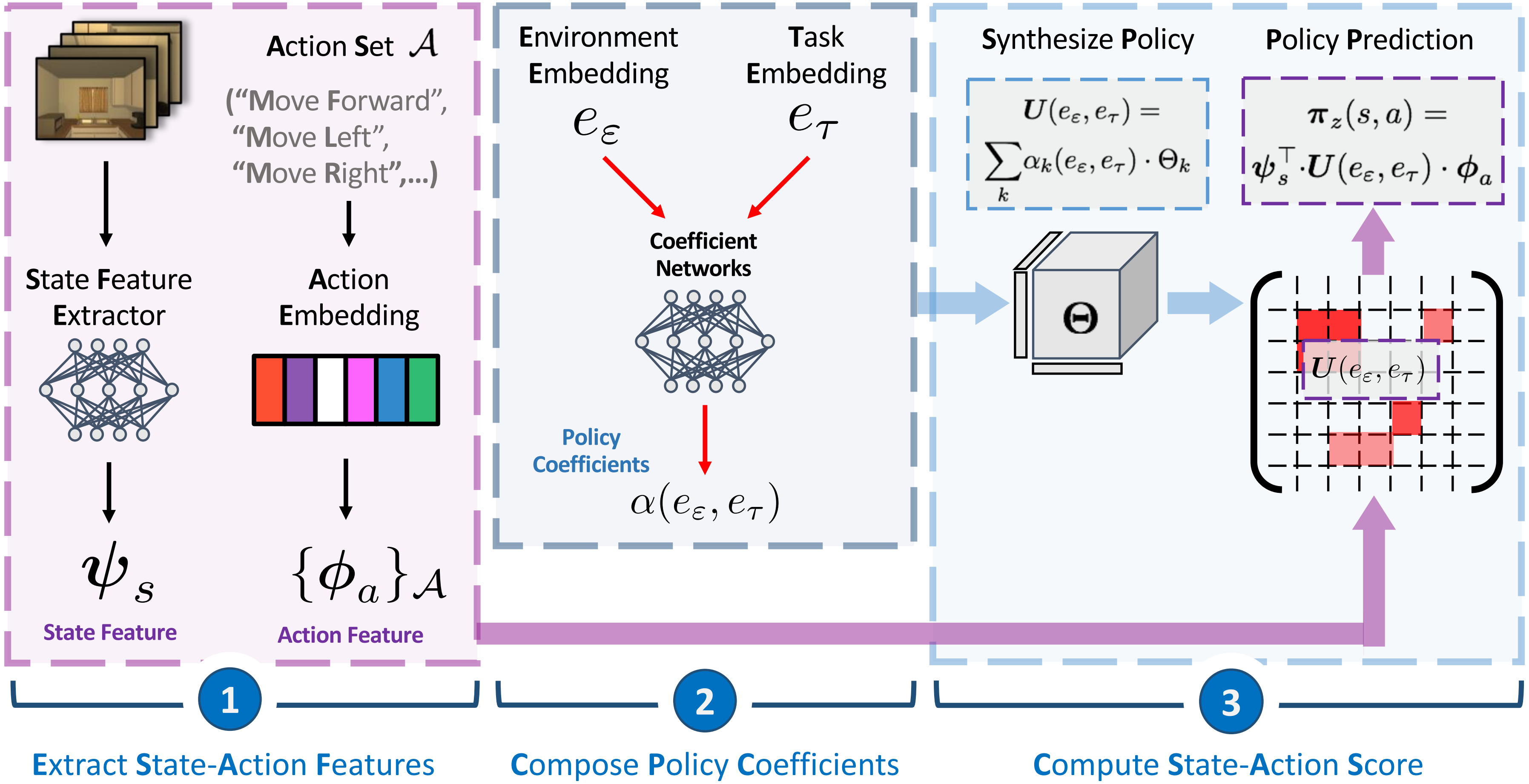
Our Method

Environment and task embeddings are learned via training on corresponding combinations.



Our Method

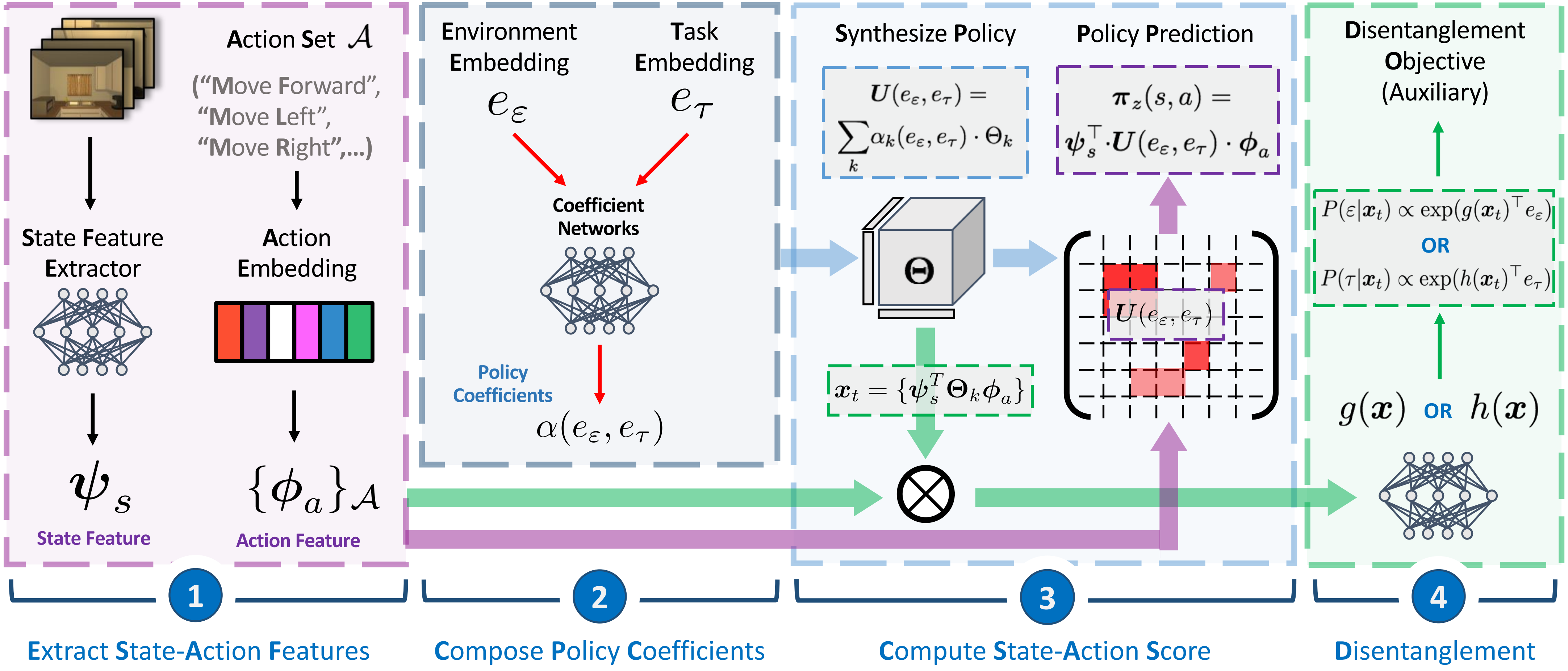
Both components are then used to compute state action compatibility score.



USC Viterbi School of Engineering

Our Method

A disentanglement objective is used as auxiliary loss term



USC Viterbi
School of Engineering

Experimental Setup

We experiment our approach on two different simulators, with **many different map** and **many tasks** (of **finding objects sequentially)**

GridWorld: 20 maps 20 tasks (144 SEEN & 256 UNSEEN)

THOR: 19 scenes 21 tasks (144 SEEN & 199 UNSEEN)

GridWorld Simulator

World

Agent's View

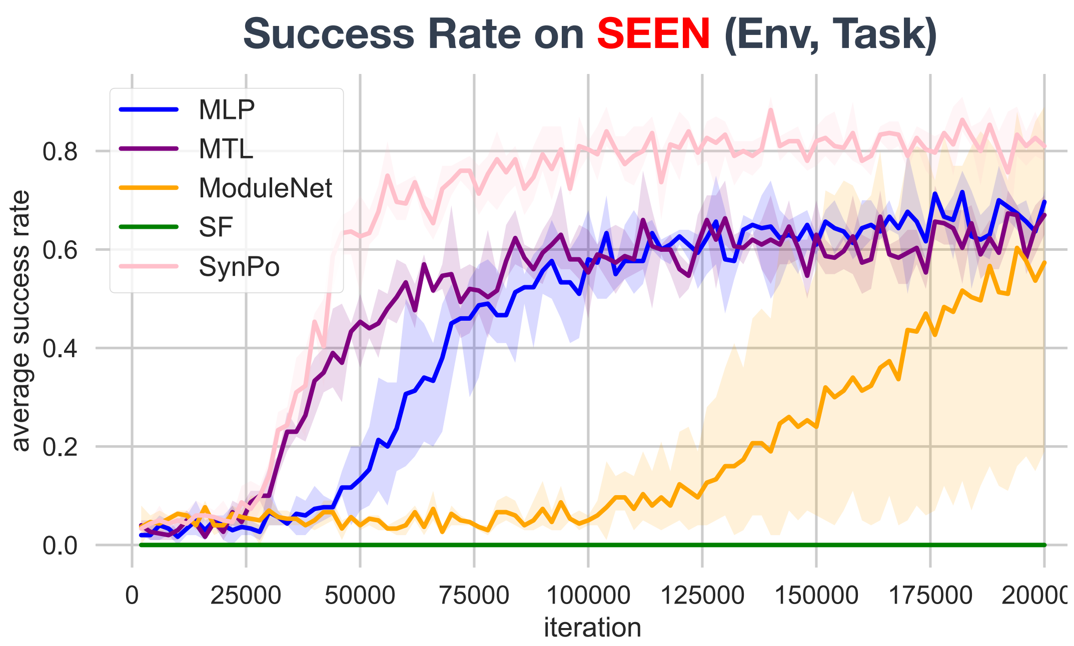
Thor [1]

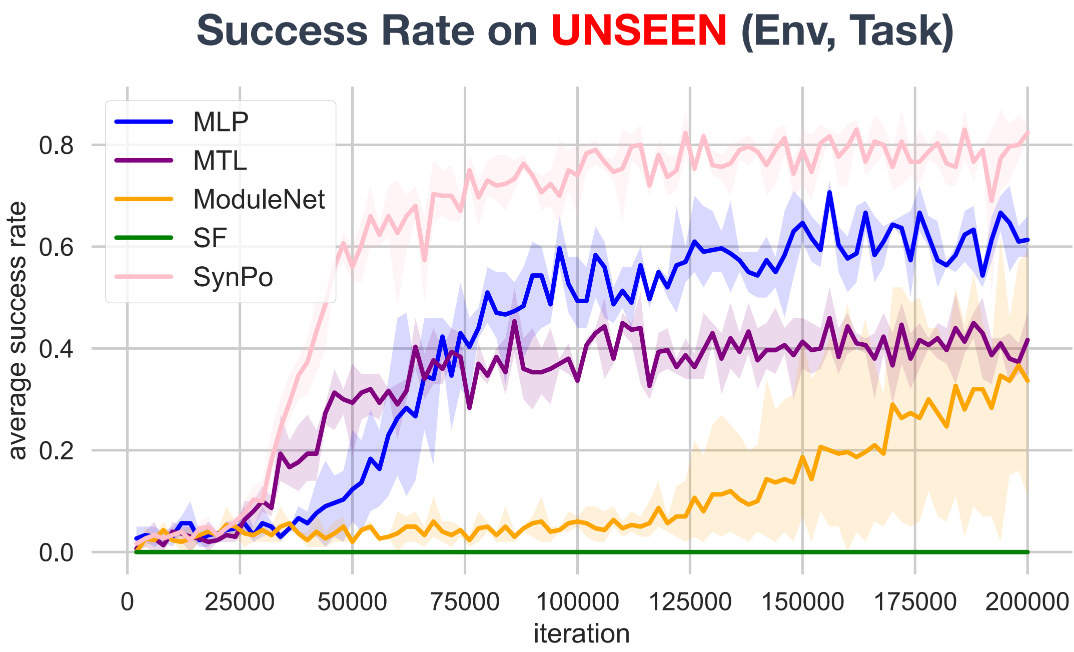
Agent's View

USC Viterbi School of Engineering

Experimental Results (Setting 1)

Performances on GridWorld





Performances on THOR[1]

Table 3: Performance of each method on THOR (SEEN/UNSEEN=144/199)

Method	ModuleNet	MLP	MTL	SYNPO
AvgSR. (SEEN)	51.5 %	47.5%	52.2%	**55.6%**
AvgSR. (UNSEEN)	14.4 %	25.8%	33.3%	**35.4%**

USC Viterbi School of Engineering

Experimental Results (Setting 2 and 3)

- **On P Set**: We train policies basis and embeddings for P's task, env
- **Setting 2**: We incrementally learn new task, environment embeddings, on **purple sets**
- **Setting 3**: We directly learn new task, environment embeddings on **Q set**

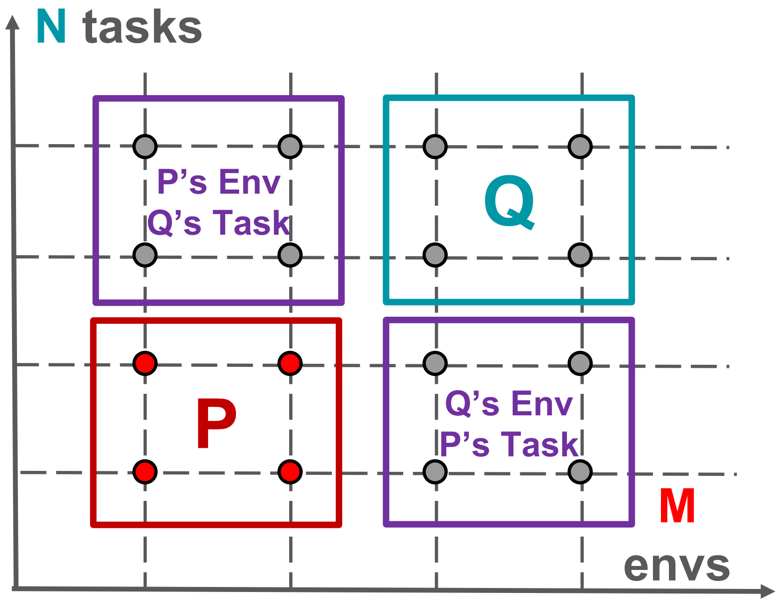


Table 2: Performance of transfer learning in the settings 2 and 3 on GRIDWORLD

Setting	Method	Cross Pair (Q's ε, P's τ)	Cross Pair (P's ε, Q's τ)	Q Pairs
Setting 2	MLP	13.8%	20.7%	6.3%
	SYNPO	**50.5%**	**21.5%**	**13.5%**
Setting 3	MLP	14.6%	18.3%	7.2%
	SYNPO	**42.7%**	**19.4%**	**12.9%**

Tencent AI Lab

NeurIPS

Thank You!

Come to our poster (#155) for more details

(Wed Dec 5th 10:45 AM – 12:45 PM @ Room 210 & 230 AB #155)

NETFLIX

USC Viterbi
School of Engineering