Knowledge Transfer through Value Function for Compositional Tasks

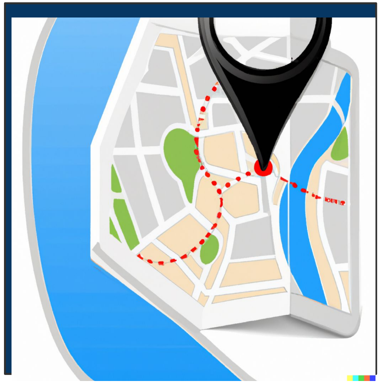
Henrique Donâncio, Matteo Leonetti, Laurent Vercouter

Motivation

Exploration is hard!

Sparse and binary outcomes

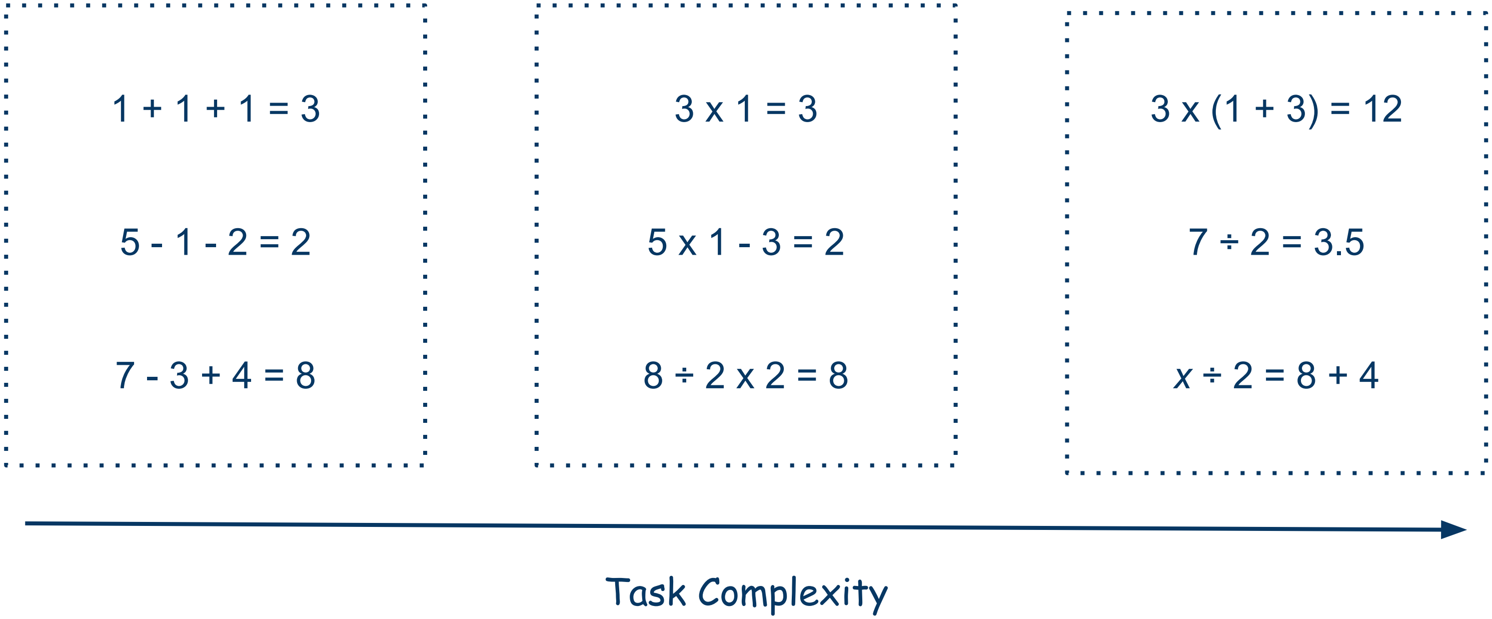
Multiple objectives



High-dimensional state spaces

All images are generated using Dalle 2 https://openai.com/dall-e-2

Curriculum Learning [Bengio et al. 2009]



1 + 1 + 1 = 3

5 - 1 - 2 = 2

7 - 3 + 4 = 8

3 x 1 = 3

5 x 1 - 3 = 2

8 ÷ 2 x 2 = 8

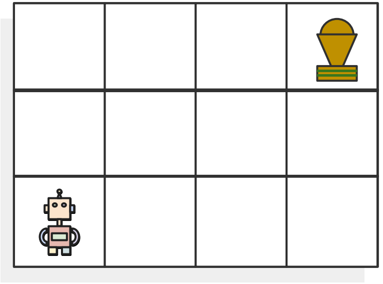
3 x (1 + 3) = 12

7 ÷ 2 = 3.5

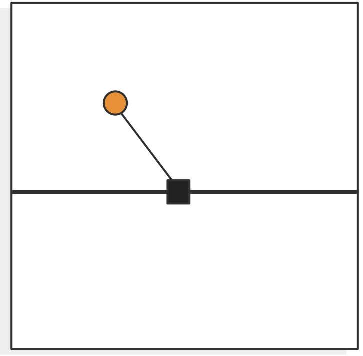
$x \div 2 = 8 + 4$

Task Complexity

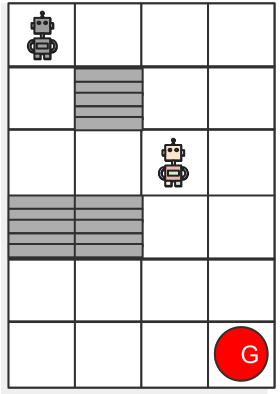
Making tasks easier: MDP degrees of freedom



Reward shaping *

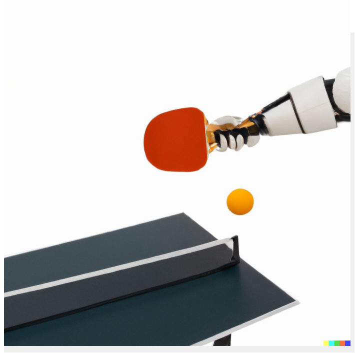


Transition probability

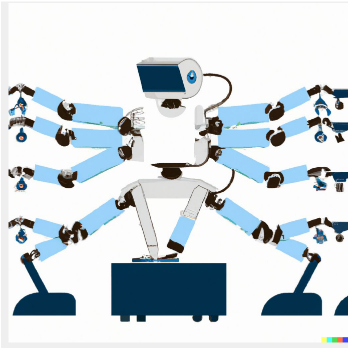


Initial state distribution

Making tasks easier: MDP degrees of freedom



States



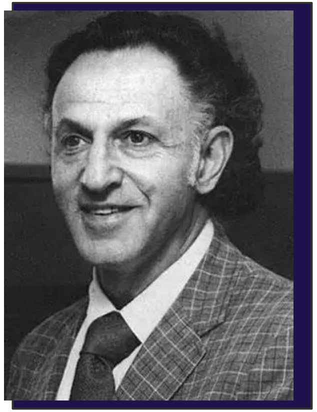
Actions

Knowledge Transfer for Compositional Representations

Q-Learning [Watkins & Dayan (1992)]

Q(s,a) How good is being in a state *s* and performing an action *a*?

$$Q(s,a) = Q(s,a) + \alpha[\overbrace{\underbrace{R(s,a) + \gamma max_{a'} Q(s',a')}_{\text{Target value}} -Q(s,a)}^{\text{TD error}}]$$



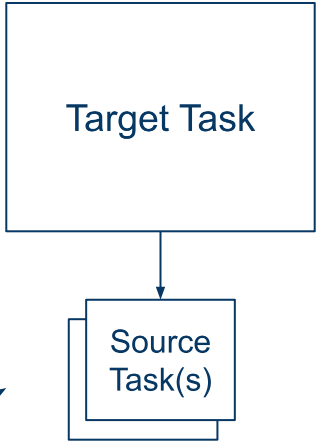
Richard Bellman

state

action

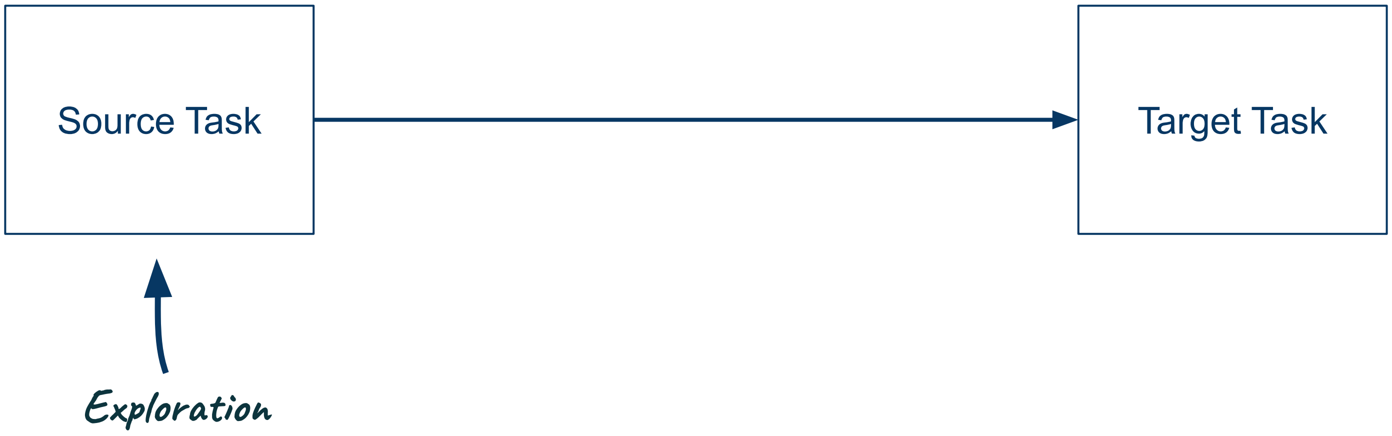
1.234	1.434	1.434	…
0.234	…		
…			

Knowledge Transfer for Compositional Representations

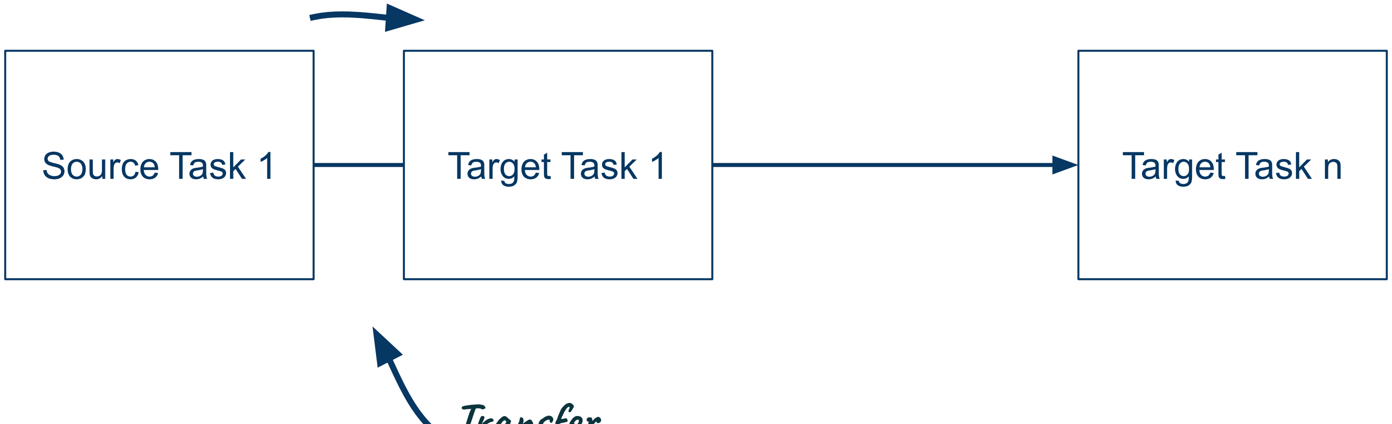


The curriculum is given!

Knowledge Transfer for Compositional Representations



Knowledge Transfer for Compositional Representations

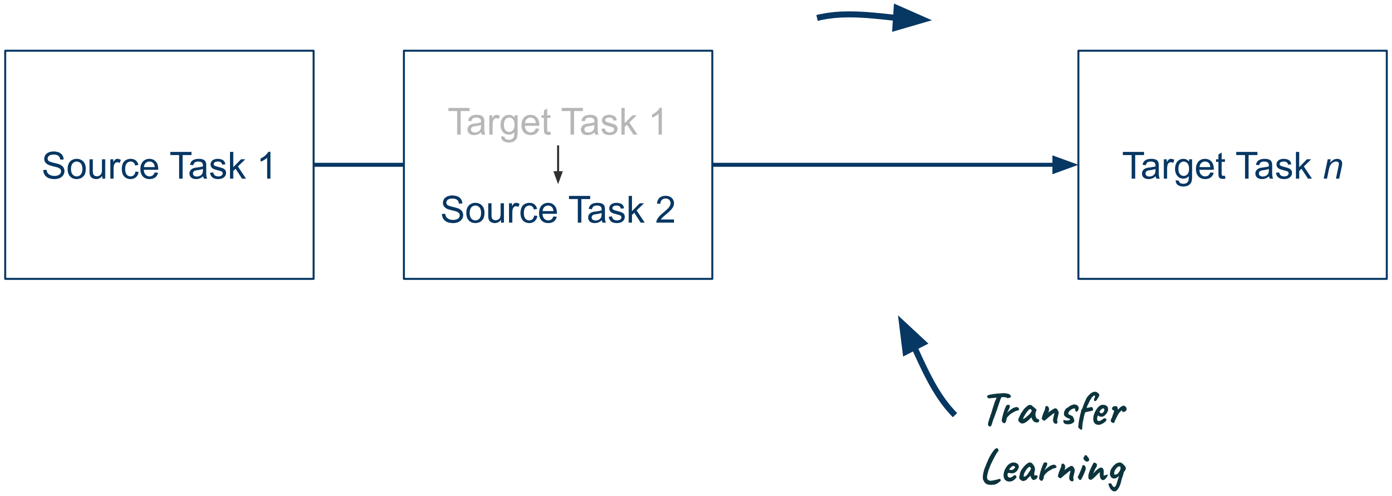


Transfer Learning

$$softmax(\{max(Q_s), max(Q_t)\})$$

$$a = \begin{cases} argmax(Q_s), softmax_{Q_s} \\ argmax(Q_t), softmax_{Q_t} \end{cases}$$

Knowledge Transfer for Compositional Representations



Source Task 1

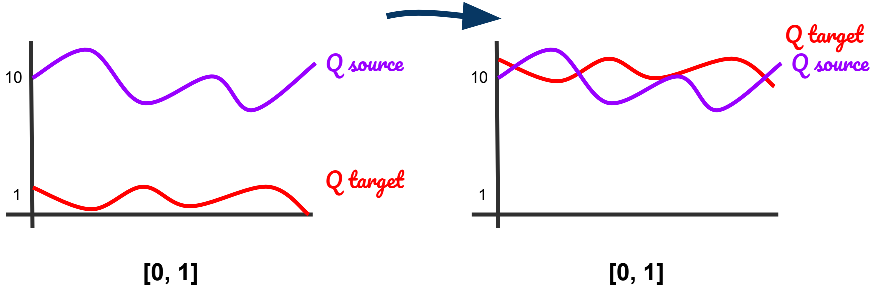
Target Task 1

Source Task 2

Target Task *n*

Transfer Learning

Knowledge Transfer for Compositional Representations

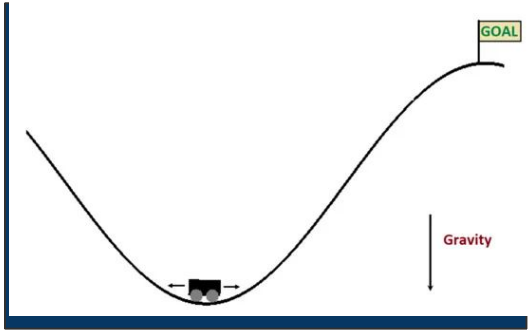


$$a = \overset{softmax(\{max(Q_s), max(Q_t)\})}{\begin{cases} argmax(Q_s), softmax_{Q_s} \\ argmax(Q_t), softmax_{Q_t} \end{cases}}$$

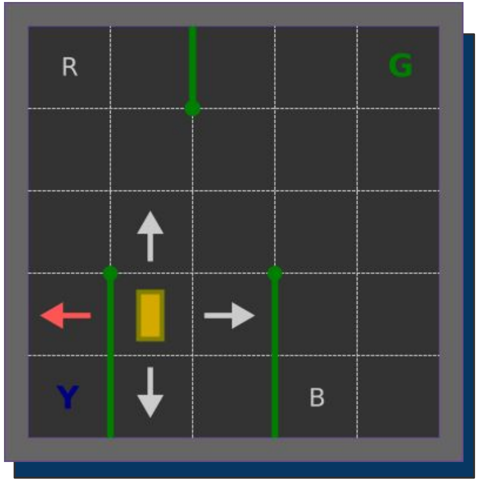
Allow target task policy play its own actions

Mitigate distributional shift

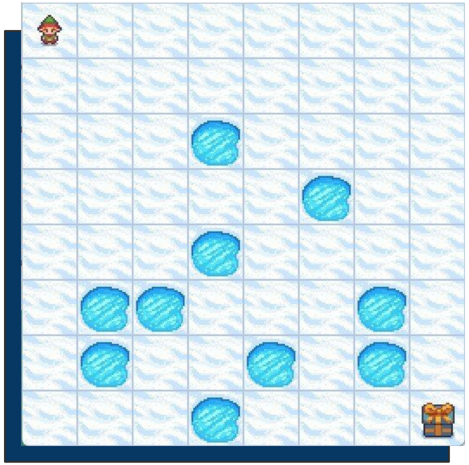
Proof-of-concept



Mountain Car

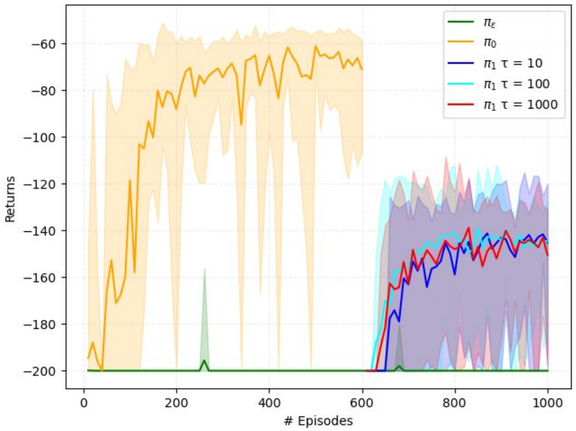


Taxi cab

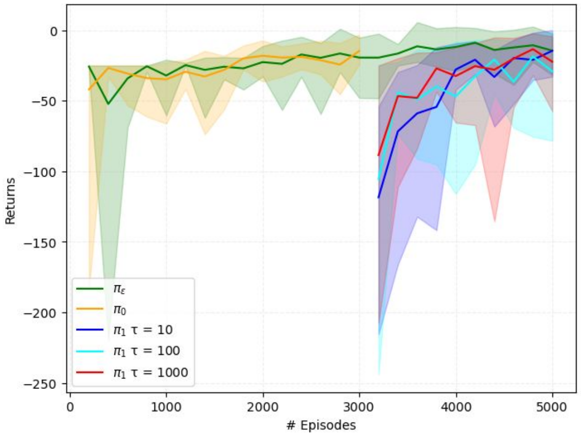


Frozen Lake

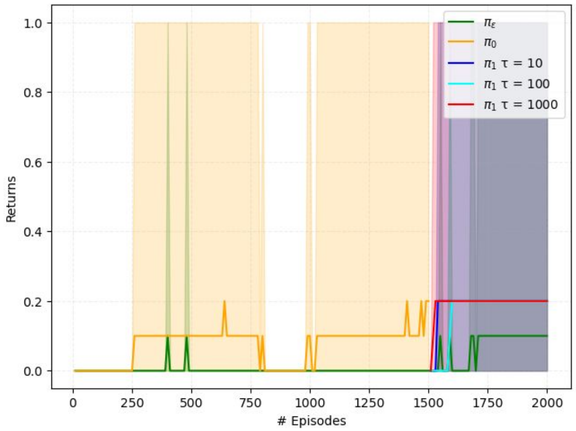
Results



Mountain Car



Taxi cab



Frozen Lake

Acknowledgments

Harald Roclawski (TU Kaiserslautern) ✧ Aloysio P. M. Saliba (UFMG) ✧ Benjamin Dewals (ULiège) ✧ Anika Theis (TU Kaiserslautern) ✧ Thomas Pirard (ULiège) ✧ Laura Sterle (TU Kaiserslautern) ✧ Thomas Krätzig (Dr. Kraetzig)

AGENCE NATIONALE DE LA RECHERCHE ANR

INSA ROUEN NORMANDIE

Water JPI

2017 JOINT CALL

The authors acknowledge FAPEMIG, Federal Ministry of Education and Research of Germany, Agence Nationale de la Recherce de France and Fonds de la Recherce Scientifique Belge, for funding this research by the project IoT.H2O (ANR-18-IC4W-0003) on the IC4Water JPI call.

References

Bellemare, Marc, et al. "Unifying count-based exploration and intrinsic motivation." Advances in neural information processing systems 29 (2016).

Tang, Haoran, et al. "#Exploration: A study of count-based exploration for deep reinforcement learning." Advances in neural information processing systems 30 (2017).

Gregor, Karol, Danilo Jimenez Rezende, and Daan Wierstra. "Variational intrinsic control." arXiv preprint arXiv:1611.07507 (2016).

Bengio, Yoshua, et al. "Curriculum learning." Proceedings of the 26th annual international conference on machine learning (2009).

Nakamoto, Mitsuhiko, et al. "Cal-QL: Calibrated Offline RL Pre-Training for Efficient Online Fine-Tuning." arXiv preprint arXiv:2303.05479 (2023).

Andrychowicz, Marcin, et al. "Hindsight experience replay." Advances in neural information processing systems 30 (2017).

Dai, Siyu, Andreas Hofmann, and Brian Williams. "Automatic curricula via expert demonstrations." arXiv preprint arXiv:2106.09159 (2021).

Watkins, Christopher JCH, and Peter Dayan. "Q-learning." Machine learning 8 (1992): 279-292.

Matthew E. Taylor, Peter Stone, and Yaxin Liu. Transfer Learning via Inter-Task Mappings for Temporal Difference Learning. Journal of Machine Learning Research, 8(1):2125–2167 (2007).