PRESENTER
Simon Rosen

Site

Background

Cooperative Multi-Agent Problems are ubiquitous in real-world applications & many involve generalising over a distribution of tasks,
- E.g. automated warehouses
- Existing solution: MARL
  - Does not scale well (sample efficiency) with increasing number of goals!

Goal-Oriented Tasks contain set of goals that agents must reach. For each task, each goal has an associated terminal reward.

World denotes a distribution of tasks that share dynamics, states, actions & non-terminal rewards. Only objectives of task (defined by goal rewards) change across task distribution

Optimal Task Gen.

World Value Function
- Goal oriented
- Extends reward function to penalise agents for reaching wrong goal
- Encodes how to reach every joint-goal (mastery)

Leverage Shared Knowledge
- How to reach goals shared across tasks
- WVF encodes this
- Transfer knowledge to new task

Zero-Shot Infer
- Given what changes between tasks (terminal rewards) and learnt WVF for a task infer new task without further learning

Scalable

Our method scales very well with the number of goals in the environment since it only needs to learn one task before it can optimally generalize over the entire task distribution.
- Size of task distribution is exponential in number of goals

The Maths

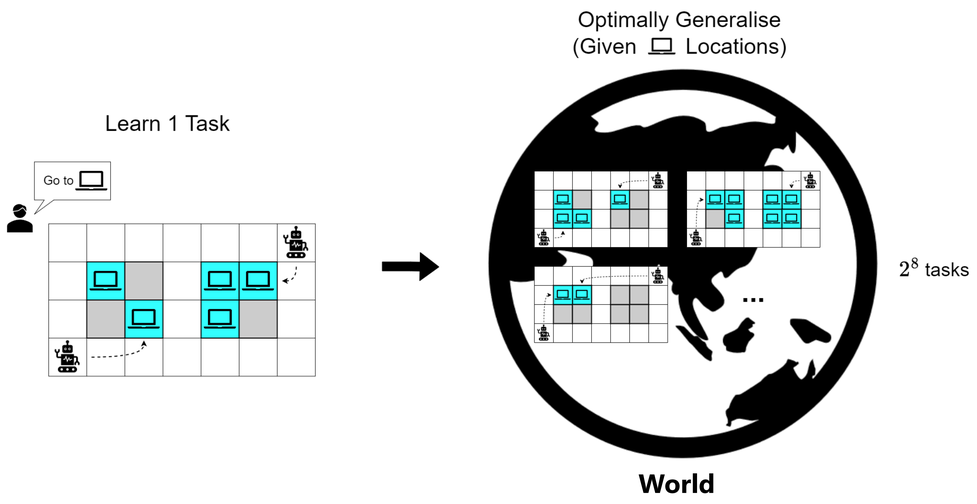
$$Q^*_{M_2}(\mathbf{s},\mathbf{a}) = \max_{\mathbf{g}\in\mathcal{G}}\{\bar{Q}^*_{M_1}(\mathbf{s},\mathbf{g},\mathbf{a}) + \sum_{i\in I}(\hat{R}^i_{M_1}(g^i) - \hat{R}^i_{M_2}(g^i))\mathbb{1}_{s^i\in\mathcal{G}^i}\}$$

- $Q^*_{M_2}(\mathbf{s},\mathbf{a})$: Inferred optimal Q-value function for task M_2
- $\bar{Q}^*_{M_1}(\mathbf{s},\mathbf{g},\mathbf{a})$: Learnt optimal WVF for task M_1
- $\hat{R}^i_{M_1}(g^i)$, $\hat{R}^i_{M_2}(g^i)$: Goal reward functions for M_1 and M_2

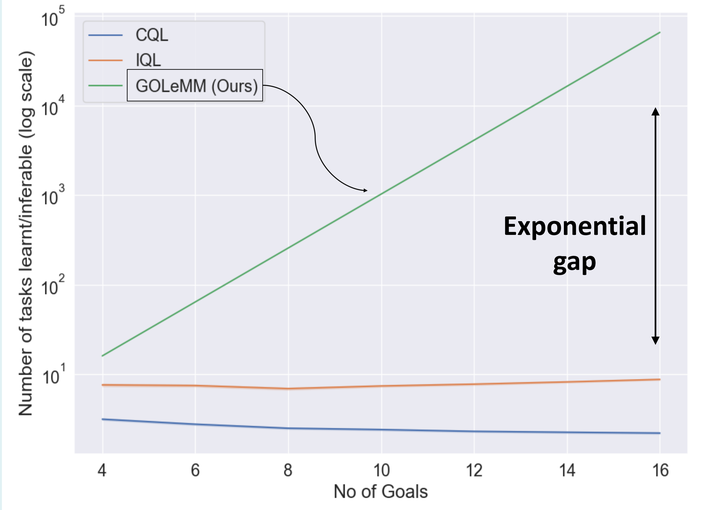
$\mathcal{G}$ - Set of joint goals I - Set of Agents

Scalable Multi-Agent RL Through Optimal Task Generalisation

Optimal Task Generalisation



Scalable



Pre-Print

SIMON ROSEN, ABDEL MFOUGOUON NJUPOUN, GERAUD NANGUE TASSE, STEVEN JAMES & BENJAMIN ROSMAN
UNIVERSITY OF THE WITWATERSRAND, RAIL LAB