A Boolean Task Algebra For Reinforcement Learning

Geraud Nangue Tasse*, Steven James and Benjamin Rosman

University of the Witwatersrand, Johannesburg, South Africa

We formalise the logical composition of tasks as a Boolean Algebra and provide a method for producing the optimal value functions of the composed tasks with no further learning.

Introduction

- We want to **combine** policies learned in previous tasks to **create new policies**.
- Build **rich behaviours** from simple ones, resulting in **combinatorial** explosion in abilities.
- But unclear how to produce new optimal policies from known ones.

Prior work [1,2] shows that value functions can be **composed** to **optimally** solve **union of tasks** and **approximately** solve the **intersection of tasks**.

We complement these results by proving **optimal composition** for the **intersection and negation of tasks** in the total-reward, absorbing-state setting, with **deterministic dynamics**.

Goal Oriented RL

We define an **extended value function (EVF)** that decouples the values for each absorbing state:

$$Q(s,g,a) = \bar{r}(s,g,a) + \int_S V^{\pi_g}(s')\rho_{(s,a)}(ds')$$

$$\bar{r}(s,g,a) = \begin{cases} N & if\ g \neq s \in G \\ r(s,a) & otherwise \end{cases}$$

Similar to DG functions [3] but uses **task rewards.**

[1] B. Van Niekerk, S. James, A. Earle and B. Rosman. Composing Value Functions in Reinforcement Learning. In ICML 2019.
[2] T. Haarnoja, V. Pong, A. Zhou, M. Dalal, P. Abbeel, and S. Levine. Composable Deep Reinforcement Learning for Robotic Manipulation.
[3] Kaelbling, L. P. Learning to achieve goals. IJCAI 1993.

Compositionality

Theorem 1: Let **M** be the set of tasks. Then **M** forms a Boolean algebra when equipped with the **or**, **and**, and **not** operators given by:

$$\ or\ = (S,A,p,r(\ or\ )), \quad where \quad r(\ or\ ) = \max\{r(\ ),r(\ )\}$$
$$\ and\ = (S,A,p,r(\ and\ )), \quad where \quad r(\ and\ ) = \min\{r(\ ),r(\ )\}$$
$$not\ = (S,A,p,r(not\ )), \quad where \quad r(not\ ) = (r_{MAX} + r_{MIN}) - r(\ )$$

where, r_{MAX} and r_{MIN} are the reward functions for the maximum and minimum tasks.

Theorem 2: Let **Q** be the set of extended value functions. Then **Q** forms a Boolean algebra when equipped with the **or**, **and**, and **not** operators given by:

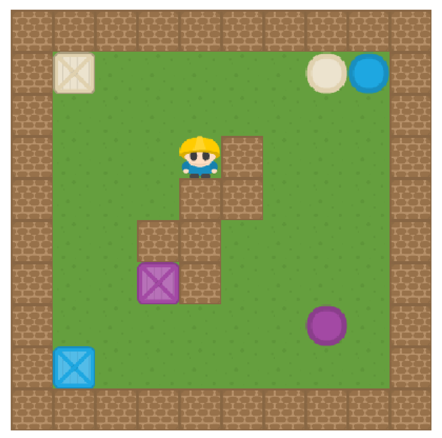
$$Q^*(\ )\ or\ Q^*(\ ) = \max\{Q^*(\ ),Q^*(\ )\}$$
$$Q^*(\ )\ and\ Q^*(\ ) = \min\{Q^*(\ ),Q^*(\ )\}$$
$$not\ Q^*(\ ) = (Q^*_{MAX} + Q^*_{MIN}) - Q^*(\ )$$

where, Q^*_{MAX} and Q^*_{MIN} are the extended value functions for the maximum and minimum tasks.

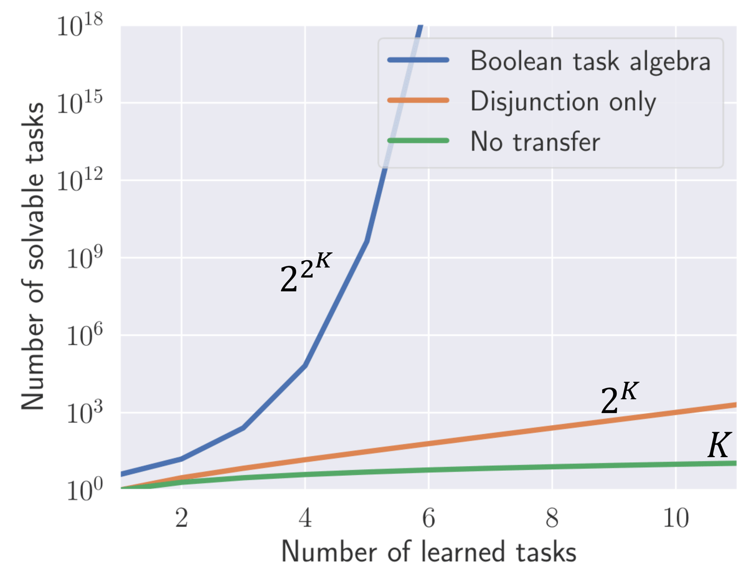
Theorem 3: The task and extended value function spaces are homomorphic.

Base Tasks and Explosion of Skills

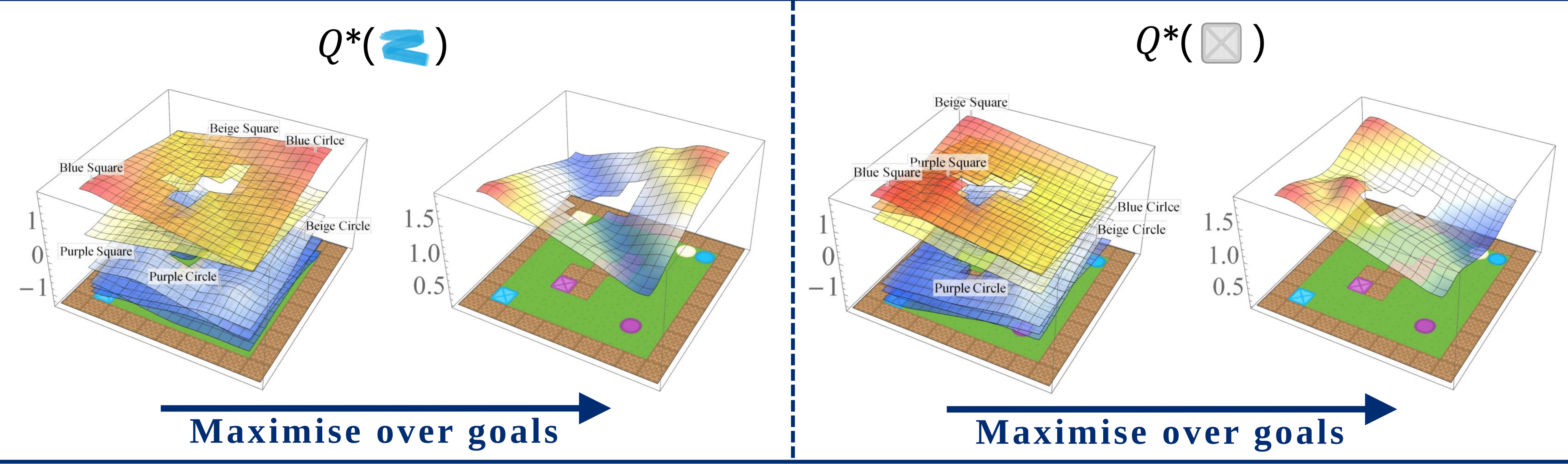
$n = |G|$ goals → 2^n total tasks
→ $[\log_2 n]$ base tasks. ($n > 1$)



	0	0	0
	0	0	1
	0	1	0
	0	1	1
	1	0	0
	1	0	1



Experiment: EVFs → VFs



Experiment (Q-Learning): Four Rooms