Self-Efficacy Update in Reinforcement Learning: Impact on Goal Selection for Q-learning Agents

Anonymous Author(s)
Affiliation
Address
email

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

Abstract

We introduce a dynamic self-efficacy learning rule and examine its impact on multi-goal selection in a grid-world. We model the Q-learning agent's self-efficacy as the integral of reward prediction errors (RPEs), allowing it to modulate the agent's expectation of achieving the best possible future outcome. Initial simulation results suggest that faster self-efficacy updates lead to higher overall reward accumulation, but with increased variability in reaching the optimal goal. These findings indicate that an optimal self-efficacy update rate, which can be learned through experience, may strike a balance between maximizing performance and maintaining stability.

1 Introduction

Multi-goal selection in reinforcement learning is challenging due to the need to balance exploration and exploitation across multiple objectives [Ecoffet et al., 2021]. Intrinsic motivation has been proposed as a mechanism to facilitate learning by integrating signals related to learning progress and competence [Barto, 2013, Colas et al., 2018]. Here, we model self-efficacy, the belief in one's ability to achieve desired outcomes [Bandura, 1997], in Q-learning agents and explore how varying self-efficacy updates impact goal selection and performance in a multi-goal grid-world environment.

2 Model

We model the *dynamics of self-efficacy* in response to feedback from the external environment and propose that self-efficacy is a dynamic attribute, continuously shaped by action outcomes. The reward prediction error (RPE) is computed from a mixture of weighted best and worst future outcomes:

$$\delta_t = R_{t+1} + \gamma \cdot (w_t \cdot \max Q(s_{t+1}, a_{t+1}) + (1 - w_t) \cdot \min Q(s_{t+1}, a_{t+1})) - Q(s_t, a_t) \quad (1)$$

The *self-efficacy belief*, w_t, scales the highest possible future expected reward contingent on action, reflecting the agent's belief that it can successfully select the best action in the immediate future (as opposed to the worst one) [Gaskett, 2003, Zorowitz et al., 2020]. RPEs serve as the critical signal for updating both the action values and the self-efficacy parameter [Li and Radulescu, 2024]:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \cdot \delta_t \quad (2)$$

$$w_{t+1} = w_t + w_{LR+} \cdot \delta_t \quad (3)$$

3 Results and Discussion

In the multi-goal grid-world environment we tested, there are two target locations with rewards: Goal 1, located at (9, 9) with a reward of 4, and Goal 2, located at (0, 9) with a reward of 3. We evaluated the impact of varying self-efficacy updates on the agent's goal selection behavior across 100 runs, each run consisting of 200 episodes. With a low self-efficacy update rate, the agent reached Goal 1 in 51.22% of runs, accumulating an average reward of 636 with a standard deviation of 23.68. For the medium self-efficacy update rate, The agent reached Goal 1 in 57.49% of runs and accumulated an average reward of 660 with a standard deviation of 27.11. With the fast self-efficacy update rate, the agent reached Goal 1 in 56.20% of runs and accumulated an average reward of 661 with a standard deviation of 32.59. Overall, the medium self-efficacy update rate provided the best balance between performance and stability, while fast updates led to increased variability in goal-selection behavior.

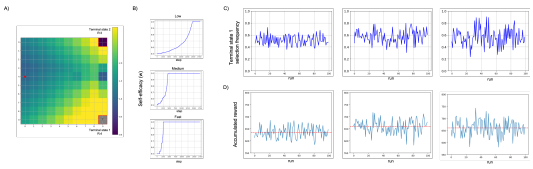

Figure 1: Impact of self-efficacy update rates on self-efficacy dynamics and goal selection behavior.

References

Albert Bandura. *Self efficacy: the exercise of control*. W. H. Freeman, New York (N.Y.), 1997. ISBN 978-0-7167-2850-4.

Andrew G. Barto. Intrinsic Motivation and Reinforcement Learning. In Gianluca Baldassarre and Marco Mirolli, editors, *Intrinsically Motivated Learning in Natural and Artificial Systems*, pages 17–47. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. ISBN 978-3-642-32374-4 978-3-642-32375-1. doi: 10.1007/978-3-642-32375-1_2. URL http://link.springer.com/10.1007/978-3-642-32375-1_2.

Cédric Colas, Pierre Fournier, Olivier Sigaud, Mohamed Chetouani, and Pierre-Yves Oudeyer. CURIOUS: Intrinsically Motivated Modular Multi-Goal Reinforcement Learning. *Proceedings of the 36th International Conference on Machine Learning 2019*, 2018. doi: 10.48550/ARXIV.1810.06284. URL https://arxiv.org/abs/1810.06284. Publisher: arXiv Version Number: 4.

Adrien Ecoffet, Joost Huizinga, Joel Lehman, Kenneth O. Stanley, and Jeff Clune. First return, then explore. *Nature*, 590(7847):580–586, February 2021. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-020-03157-9. URL https://www.nature.com/articles/s41586-020-03157-9.

Chris Gaskett. Reinforcement learning under circumstances beyond its control. *In Proceedings of the international conference on computational intelligence, robotics and autonomous systems*, 2003.

Jing Li and Angela Radulescu. Dynamic self-efficacy as a computational mechanism of mania emergence. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 46, 2024.

Samuel Zorowitz, Ida Momennejad, and Nathaniel D. Daw. Anxiety, Avoidance, and Sequential Evaluation. *Computational Psychiatry*, 4(0):1, 2020. ISSN 2379-6227.