

1 Reinforcement Learning

1.1 Reward function

1.2 Types of reward signals

1.2.1 Sparse Reward

Curse of sparse reward.

许多有趣的问题最自然地被描述为稀疏的奖励信号。在稀疏环境中寻找最优策略通常是不可行的，因为依赖非定向探索 (undirected exploration) 的通用算法需要“大得令人望而却步”的训练步数。

1.2.2 Dense Reward

促进学习 (Facilitating Learning)：稀疏奖励可以通过基于势能的奖励塑造 (potential-based reward shaping) 转换为密集奖励。这种转换旨在极大地促进学习过程。

连续逼近 (Successive Approximation)：历史悠久的行为科学技术——塑造 (shaping, 也称为“连续逼近”) ——通过分解造成稀疏奖励难以学习的行为上的不连续性 (discontinuities)，从本质上创建了一个密集的学习信号。例如，斯金纳 (Skinner) 就曾利用基于角度和位置的奖励 (密集信号) 来引导一只鸽子到达目标区域，最终的稀疏奖励在那里等待着它。

1.2.3 Reward Shaping

奖励塑造 (Reward Shaping)：可以通过势函数 (potential-based) 的转换，将稀疏奖励转化为密集奖励 (dense reward)，从而极大地促进学习过程。这种变换必须保证最优策略保持不变。

自然奖励信号 (Natural Reward Signals)

1. 进化奖励信号：生存与适应度

Evolutionary Reward Signals: Survival and Fitness

2 训练动力学