



# Reinforcement Learning: An Introduction

强化学习导论第二版疑问

作者：吕昀璘

组织：UESTC

时间：March 24, 2020



*Victory won't come to us unless we go to it. — M. Moore*

# 目录

<b>1</b>	<b>介绍</b>	<b>1</b>
1.1	强化学习	1
1.2	例子	1
1.3	强化学习的要素	1
1.4	局限和范围	1
1.5	拓展例子：井字游戏	1
1.6	总结	1
1.7	强化学习的早期历史	1
	<b>第一部分 表格解决方法</b>	<b>2</b>
<b>2</b>	<b>多臂赌博机</b>	<b>3</b>
2.1	$k$ 臂赌博机问题	3
2.2	动作值方法	3
2.3	10 臂试验	3
2.4	渐增实现	3
2.5	非平稳问题	3
2.6	乐观初始值	3
2.7	置信上限动作选择	3
2.8	梯度赌博机算法	3
2.9	关联搜索（上下文赌博机）	3
2.10	总结	3
<b>3</b>	<b>有限马尔可夫决策过程</b>	<b>4</b>
3.1	Agent-环境接口	4

# 第一章 介绍

---

## 1.1 强化学习

## 1.2 例子

## 1.3 强化学习的要素

## 1.4 局限和范围

## 1.5 拓展例子：井字游戏

## 1.6 总结

## 1.7 强化学习的早期历史

## 第一部分

# 表格解决方法

## 第二章 多臂赌博机

---

### 2.1 $k$ 臂赌博机问题

### 2.2 动作值方法

### 2.3 10 臂试验

### 2.4 渐增实现

### 2.5 非平稳问题

$$(1 - \alpha)^n + \alpha \sum_{i=1}^n (1 - \alpha)^{n-i} \quad (2.1)$$

$$= (1 - \alpha)^n + \alpha \frac{1 - (1 - \alpha)^n}{1 - (1 - \alpha)} \quad (2.2)$$

$$= (1 - \alpha)^n + 1 - (1 - \alpha)^n \quad (2.3)$$

$$= 1 \quad (2.4)$$

### 2.6 乐观初始值

### 2.7 置信上限动作选择

### 2.8 梯度赌博机算法

### 2.9 关联搜索（上下文赌博机）

### 2.10 总结

## 第三章 有限马尔可夫决策过程

---

### 3.1 Agent-环境接口

### 3.2 目标和奖励

### 3.3 回报和 episode

### 3.4 回合和连续任务的统一符号

### 3.5 策略和值函数

### 3.6 最优策略和最优值函数

### 3.7 最优和近似

### 3.8 总结

## 第四章 动态规划

