很强的相互作用，并对这两个学科做出了重大的贡献。在机器学习的所有不同形式中，强化学习是和人类以及其他动物最接近的一种学习方式，并且它的许多核心算法最初也受到生物学习系统的启发。同时强化学习也由此做出了自己的成果：它给出的动物学习心理学模型能够更好地适用于一些经验数据，它也提出了一个大脑收益机制部分的重要模型。本书主要讲述了工程和人工智能范畴内的强化学习思想，与心理学和神经科学相关的内容将在第 14 和 15 章介绍。

最后，强化学习也投身到回归简单普适原则的人工智能大趋势中，从 20 世纪 60 年代末期以来，大量的人工智能研究者认为不存在什么普适的原则，智力源于针对大量特定目的产生的技巧、过程和启发式方法。有时人们会说，如果我们能把足够多的有关信息放到一台机器上，比如百万级或十亿级的信息，那么机器就会变得智能。基于一般原则的方法，比如搜索或学习，被定性为“弱方法”；而基于知识的方法则被称为“强方法”。这种观点在今天仍然很流行，但却没有那么强的主导地位了。我们不能仅作少量的尝试和努力就得出没有普适性方法的结论。在现代人工智能领域已经做了许多研究以寻找学习、搜索和决策的普适原则，尝试引入大量的领域知识。虽然目前还不清楚会在这个方向上做到什么程度，但强化学习研究无疑在追求更简单的人工智能普适原则。

1.2 示例

下面我们通过一些案例和应用来理解强化学习：

- 国际象棋大师走一步棋。这个选择是通过反复计算对手可能的策略和对特定局面位置及走棋动作的直观判断做出的。
- 自适应石油控制器实时调整石油提炼过程参数。该控制器以特定的边际成本为基础，权衡优化产品的收益率/成本/质量，不需要严格遵守工程师的初始设置。
- 一只羚羊幼崽出生后数分钟挣扎着站起来。半小时后，它能够以每小时 20 英里的速度奔跑。
- 一个移动机器人决定它是进入一个新房间收集更多垃圾还是返回充电站充电。它的决定将基于当前电量，以及它过去走到充电站的难易程度。
- 菲尔准备早餐。仔细看一下，即使是这个看似平凡的日常活动也揭示了一个复杂的网络，其中包含特定条件下的行为和连锁的目标关系：走近橱柜，打开橱柜，选择一种麦片粥盒子，然后伸手去拿，抓住，取出盒子。而取出碗、勺子和牛奶、壶这些动作则是其他复杂的需要调整的交互式的行为序列。这些行为的每一步都需要我