

---

# 计算机科学188：人工智能导论

## 2024年春季

---

Note 7

作者（其他所有注释）：尼基尔·夏尔马

作者（贝叶斯网络注释）：乔希·胡格和杰基·梁，由王蕾吉娜编辑

作者（逻辑注释）：亨利·朱，由考佩林编辑

学分（机器学习与逻辑注释）：部分章节改编自教材《人工智能：一种现代方法》。

最后更新时间：2023年8月26日

## 基于知识的智能体

想象一个充满熔岩的危险世界，唯一的避难所是远处的绿洲。我们希望我们的智能体能够从当前位置安全地导航到绿洲。

在强化学习中，我们假设我们能给出的唯一指导是一个奖励函数，它会试图将智能体推向正确的方向，就像一场“热或冷”的游戏。随着智能体探索并收集更多关于世界的观察结果，它逐渐学会将一些行动与积极的未来奖励联系起来，而将其他行动与不良的、滚烫的死亡联系起来。通过这种方式，它可能学会从世界中识别某些线索并据此行动。例如，如果它感觉到空气变热，就应该转向另一个方向。

然而，我们可以考虑另一种策略。相反，我们告诉智能体一些关于世界的事实，并让它根据手头的信息来推断该做什么。如果我们告诉智能体，在熔岩坑周围空气会变得炎热且朦胧，而在水体周围空气会清新洁净，那么它就可以根据对大气的读数合理推断出景观中的哪些区域危险或安全。这种类型的智能体被称为基于知识的智能体。这样的智能体维护一个知识库，它是一组逻辑语句的集合，这些语句编码了我们告诉智能体的内容以及它所观察到的内容。智能体还能够进行逻辑推理以得出新的结论。

## 逻辑语言

与任何其他语言一样，逻辑语句是用一种特殊的语法编写的。每个逻辑语句都是关于一个可能为真或可能不为真的世界的命题代码。例如，“地面是熔岩”这句话在我们智能体的世界中可能为真，但在我们的世界中可能不为真。我们可以通过用逻辑连接词将更简单的语句串在一起，来构造复杂的语句，比如“从大C可以看到整个校园，徒步旅行是从学习中解脱出来的健康方式”。这种语言中有五个逻辑连接词：

- $\neg$ ，而非：当且仅当 (iff)  $P$  为假时， $\neg P$  为真。原子语句  $P$  和  $\neg P$  被称为文字。

- $\wedge$ ，且：当且仅当  $A$  为真且  $B$  为真时， $A \wedge B$  为真。“且”语句被称为合取式，其组成命题称为合取项。

- $\vee$ ，或者说：当  $A$  为真或者  $B$  为真时， $A \vee B$  为真。一个“或”语句被称为析取，其组成命题称为析取项。
- $\Rightarrow$ ，蕴含：除非  $A$  为真且  $B$  为假，否则  $A \Rightarrow B$  为真。
- $\Leftrightarrow$ ，双条件：当  $A$  和  $B$  都为真或者都为假时， $A \Leftrightarrow B$  为真。

$P$	$Q$	$\neg P$	$P \wedge Q$	$P \vee Q$	$P \Rightarrow Q$	$P \Leftrightarrow Q$
false	false	true	false	false	true	true
false	true	true	false	true	true	false
true	false	false	false	true	false	false
true	true	false	true	true	true	true

图7.8 五个逻辑连接词的真值表。例如，要使用该表计算当  $P$  为真且  $Q$  为假时  $P \vee Q$  的值，首先在左边找到  $P$  为真且  $Q$  为假的行（第三行）。然后在该行的  $P \vee Q$  列下查看结果：真。