Imitation Learning: A Survey of Learning Methods 阅读报告

韩馥光

南京大学人工智能学院

December 4, 2020

南京大学 人工智能学院
School of Artificial Intelligence, Nanjing University

目录

目录

引言

- 近年来，能够模仿人类行为的智能体的需求大大增加
  - 自动驾驶，辅助机器人和人机交互
- 复杂的应用中可能出现的场景数量太大，无法通过显式编程覆盖的情况，需要能够处理看不见的场景
  - 由专家提供的先验知识比从头开始寻找解决方案更为有效和高效
  - 为每个任务单独设计奖励是非常困难的

引言

- 一种更自然，更直观的方法是针对需要学习者模仿的所需行为进行演示
- 模仿学习的工作原理是：
  - 提取专家行为和周围环境的信息
  - 学习环境信息与演示行为之间的映射

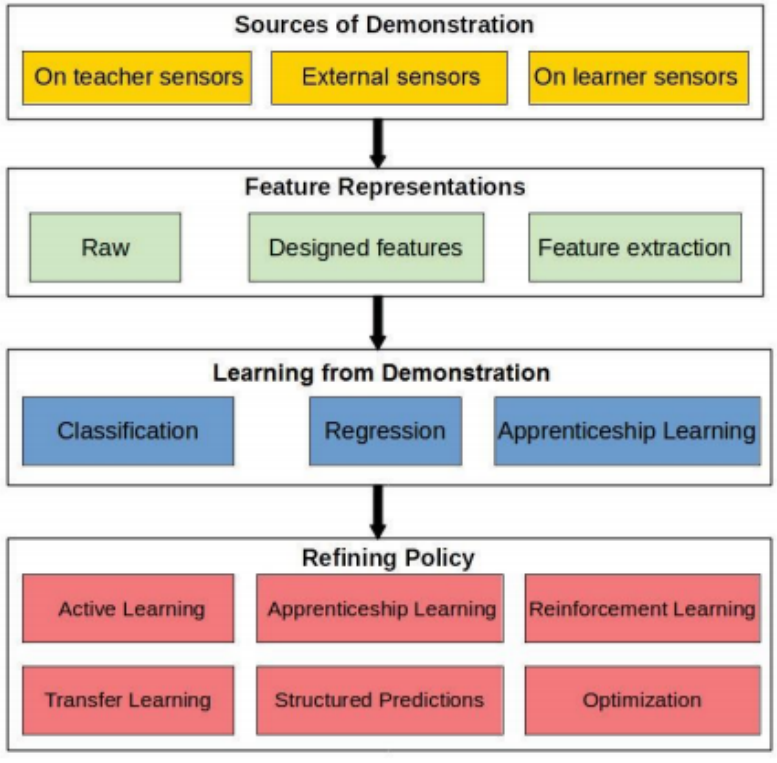

Figure: 模仿学习流程图

面临的挑战

- 传感器捕获的数据有噪声或错误
- teacher 和 learner 的匹配问题
- 可观察性问题
- 计算能力和内存的限制
- 泛化性问题

目录

问题形式化

定义

- policy：是将状态映射到行动的函数
- demonstration:(x, y)，x 是描述状态的特征向量，y 是演示者执行的动作
- experience:(s, a, r, s')，s 是状态，a 是在状态 s 采取的动作，r 是执行动作 a 的获得的奖赏，s' 是进行该操作产生的新状态
- non-stationary policy: 在学习策略参数时使用时间 t 的策略，即该策略考虑了时间这一参数
- stationary policy：在学习策略的过程中忽略时间这一参数的策略

除了奖赏函数外，从演示中学习和从经验中学习的学习参数相似，将两种方法进行结合比较常见
stationary policy 的一个优势是能够学习范围很大或未知的任务

目录

处理演示

挑战

- 之前介绍的匹配问题
- 演示不完整或不准确

原始特征

如果原始特征传达了足够的信息并且具有适当数量的维度，则无需进行进一步处理即可适合学习

例如一些 2D 游戏，可以直接将整个屏幕的图像作为输入而无需提取任何特征

手动设计特征

- 需要一些专业知识
- 在计算机视觉领域中很受欢迎
  - 特征从数值域到二进制图块的转换可显著改善学习

特征提取

- 主成分分析 (PCA)
  - PCA 可以用较低的维度表示状态，已被高自由度机器人广泛使用
- 深度学习方法
  - 深度学习方法也可以用于在没有专家知识的情况下提取特征，这些方法在从高维数据自动学习特征中取得了成功，如 DQN 等

自动提取的功能的优点是可以最大限度地减少训练智能体所需的特定任务的知识。这能构建更通用的学习方法，这些方法可用于通过最少的调整直接从演示中学习各种任务。

目录

直接模仿

- 分类
  - 分类方法适用于学习者的行动可以被视为为一些离散的类别
- 回归
  - 回归方法用于学习连续空间中的动作

直接模仿学习有一定的局限性，局限性归因于两个主要因素：演示错误、泛化性不强。

目录

强化学习

- 如果从演示中已经学到了策略，那么可以应用强化学习来微调参数。提供正面或负面的例子来训练策略有助于通过减少可用的搜索空间来加快学习
- 将强化学习应用于已经初步学习的策略而不是随机策略，还可以减小策略收敛到局部最小值的风险
- 强化学习对于演示中未包含的情况也可以进行学习

优化

- 定义：给定反映代理性能的成本函数 f：$A \rightarrow R$，其中 A 是一组输入参数，R 是一组实数，优化方法旨在查找输入参数 x_0 最小化成本函数，使得 $f(x_0) \leq f(x) \forall, x \in A$
- 与强化学习类似，优化技术可用于通过从随机解开始并迭代改进以优化适应度函数来找到问题的解
- 进化算法是比较流行的优化算法，已广泛用于寻找机器人任务的运动轨迹
- 进化算法可以与模仿学习结合，以改善通过演示学习的轨迹或加快优化过程

迁移学习

- 迁移学习使用某一任务或领域的知识来加强对另一任务的学习
- 迁移学习与模仿学习的结合和机器人应用有关，因为获取样本既困难又昂贵，利用我们已经学习的一项任务的知识可能是效率高并且有效的

逆向强化学习

逆向强化学习（IRL）使用训练样本来学习奖赏函数，并使用它来改进训练后的模型

主动学习

- 主动学习能够向专家查询对给定状态的最佳响应，并使用这些主动样本来改进其策略
- 主动学习是使模型适应原始训练样本未包含的情况的有效方法
- 主动学习是使用置信度估计来识别模型中需要改进的部分，决定何时需要向专家查询

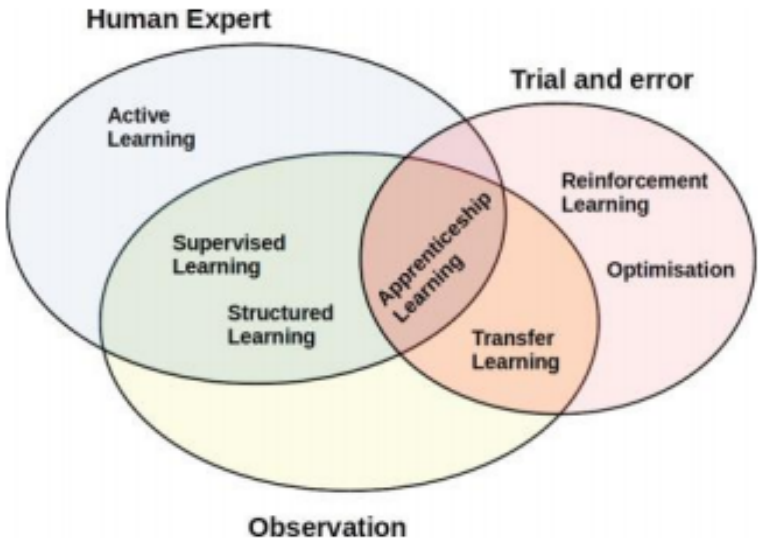

Figure: 不同来源的学习方法

目录

多智能体模仿学习

尽管缺乏研究，但是模仿学习仍然很适合应用于多智能体领域：

- 在多智能体环境中，可以从演示中学习，因为可以在目标相似的智能体之间传递知识
- 智能体在一些任务中需要以人类角度来看比较切合实际的方式进行互动，在这些任务中模仿学习是有帮助的

目录

评估

通常在专门的机器人或模拟器上进行实验。这使与其他方法的比较变得困难，原因有两个：

- 模仿学习设计的算法通常专用于特定的任务和环境
- 由于要再现实验结果需要特殊的硬件或软件，而其他研究人员可能无法使用

评估方式

- 定量评估
  - 误差计算
  - 量化分数
- 定性评估
  - 测试受过训练的智能体与人类玩家的相似程度

目录

未来方向

- General feature representations
- General task learning
- Benchmarking
- Multi-agent imitation
- Agent memory

Thank you

Thank you for listening!