Research Notes

Tao Zhu
zhutaoer@outlook.com

创建于: 2025 年 12 月 24 日

更新于: 2025 年 12 月 29 日

目录

1 Direct Data Driven Control

定义 1.1: 数据驱动控制

制器的设计不包含受控过程数学模型信息，仅利用受控系统的输入输出数据以及经过数据处理而得到的知识来设计控制器。

表 1: 数据驱动控制的典型类别

类别	描述
基于模型的间接数据驱动控制	通过数据辨识建立受控对象模型，然后基于所辨识的模型设计控制器。
直接数据驱动控制	直接利用输入输出数据设计控制器，无需辨识受控对象模型。
基于机器学习的数据驱动控制	利用机器学习方法（如强化学习、神经网络）从数据中学习控制策略。

主要类型分: 离线-在线两类

三种重要类型

- MFAC（无模型自适应控制）
- Koopman 控制
- 神经网络/深度强化学习控制

被控对象模型的作用

- 描述系统行为动态，分析系统特性（响应、超调）
- 线性模型和传递函数可以用来进行求解
- 通过给定输入求输出
- 预测、评估和决策

表 2: 数据驱动控制的典型方法示例（离线/在线）

	说明或示例
SPSA（同时扰动随机近似)	一种随机优化方法，可用于无模型或仅基于数据的参数调优。
虚拟参考反馈整定 (VRFT)	直接数据驱动的控制器整定方法，通过虚拟参考实现闭环性能匹配。
Lazy learning / 局部学习	基于局部数据集的非参数学习方法，用于在线/增量学习控制策略。
无模型自适应控制 / PID	基于误差信号进行在线参数调整的经典方法。
神经网络控制 / 深度强化学习	使用神经网络或强化学习从数据中直接学习控制策略，适合复杂/非线性系统。
Koopman 方法	通过选择合适基函数将非线性系统线性化，便于利用线性工具进行控制设计。
PINN（物理启发神经网络）	将物理守恒律或偏微分方程融入神经网络训练，提高数据效率和物理一致性。
数据驱动的 MPC	直接利用历史数据或学习到的模型构建在线预测控制器（MPC）。
线性模型假设辨识 + 最优控制	先通过数据辨识得到线性近似模型，再基于该模型设计最优控制器。

使用模型时存在的挑战

- 第一性原理不准确，需要良好的数学/物理方法建模
- 模型难以建立、建模不准确
- 建立的模型十分复杂，难以使用，解析解难以获得，数值解也难以获得；
- 系统内部和环境中的扰动影响如何考虑
- 使用模型做预报：针对确定性随机系统（混沌系统）长期预报是困难的，难以直接利用其模型
- 复杂系统（建模十分复杂，相互耦合效应不明确）

1.1 阅读:Formulas for Data-Driven Control: Stabilization, Optimality, and Robustness

Authors:Claudio De Persis and Pietro Tesi

1.2 阅读：Verhoek 博士论文

笔记 1.1: Direct Data-Driven Control 的优势

- There is no bias introduced due to incorrect model selection or unmodeled dynamics. All the behavior reflected in the data is used for control, and we only need to be robust against the noise in the data.
- The question of finding the best model for controller design becomes redundant in the direct approach because there is no model.
- Stabilization problems for unstable systems are often easier via the direct paradigm, while identifying unstable systems is often difficult.

笔记 1.2: The influence of the behavior approach(Polderman and Willems 1997) for direct data-driven LTI control(Markovsky and Dorfler 2021)

-> there is no unifying framework that allows for providing rigorous guarantees of both stability and performance in a direct data-driven setting.

2 彩色信息框示例

定义 2.1: 具身马尔可夫决策过程（EMDP）

一个具身马尔可夫决策过程定义为五元组 $\langle \mathcal{S}, \mathcal{A}, T, R, \gamma \rangle$，其中：

- $\mathcal{S} = \mathbb{R}^2 \times [0, 2\pi)$ 为状态空间（位置 + 朝向）；
- $\mathcal{A} = \{\text{前进}, \text{左转}, \text{右转}\}$ 为离散动作空间；
- $T(s' \mid s, a)$ 为状态转移函数；
- $R(s, a) = -\|s_{\text{pos}} - s_{\text{goal}}\|$ 为稀疏奖励（负欧氏距离）；
- $\gamma \in (0, 1)$ 为折扣因子。

假设 2.1: 可观测性

智能体在每一步都能精确观测当前状态 $s_t \in \mathcal{S}$。

假设 2.2: 确定性动力学

状态转移是确定性的，即 $s_{t+1} = f(s_t, a_t)$，无环境噪声。

该算法的正确性由以下定理保证。

引理 2.1: 贝尔曼最优性

值函数 V^* 是贝尔曼最优算子 $\mathcal{T}$ 的唯一不动点，即 $V^* = \mathcal{T} V^*$，其中

$$\mathcal{T} V(s) := \max_{a \in \mathcal{A}} \left[R(s, a) + \gamma \sum_{s'} T(s' \mid s, a) V(s') \right].$$

定理 2.1: 值迭代收敛性

值迭代算法生成的序列 $\{V_k\}$ 以指数速率收敛到 V^*，即

$$\|V_k - V^*\|_\infty \le \frac{2\gamma^k}{1 - \gamma} \|V_1 - V_0\|_\infty.$$

证明. 由于状态空间 $\mathcal{S}$ 在假设下为有限（或可离散化），且 $\gamma < 1$，贝尔曼算子 $\mathcal{T}$ 是 γ-压缩映射。由巴拿赫不动点定理，迭代 $V_{k+1} = \mathcal{T} V_k$ 收敛到唯一不动点 V^*，且误差界如上所述。 □

例子 2.1: 2D 网格世界导航

考虑一个 5×5 的网格世界，机器人从左下角 $(0,0)$ 出发，目标为右上角 $(4,4)$。状态 $s=(x,y,\theta)$，其中 $\theta\in\{0,\pi/2,\pi,3\pi/2\}$。动作 $\mathcal{A}=\{前进,左转,右转\}$ 控制朝向与位置。在 **??** 下，值迭代可精确计算到达目标的最短路径。