

# VEOS 系统评估

忻斌健

2019 年 9 月 3 日

## 测试条件

- 固定测试场景
  - 静止到匀加速再匀减速到停止
- 固定工况
  - 不开空调（减少空调能耗干扰）
  - 往返路线（减少地形差异干扰）
- 不可控的观测噪声：
  - 地形
  - 压缩机
  - 电池 SOC
  - 大灯
  - tbox
- 测量驾驶风格
  - 纵向控制问题中，特定工况下油门踏板（和刹车踏板）的使用情况
- 通过独立的 UDP 数据记录交叉验证测量和性能
- 总共实验约 1500 次

## 驾驶风格

### 无 AI 和带 AI 的基准驾驶风格比较

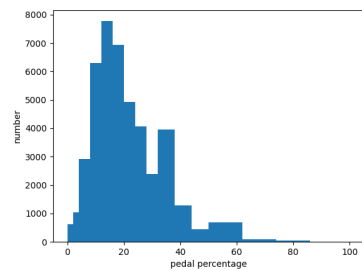


图 1.1 无 AI 的基准风格分布

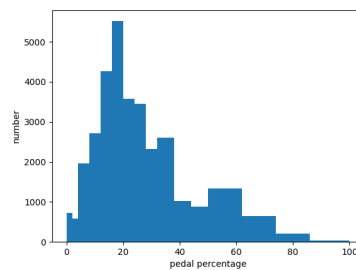


图 1.2 带 AI 的基准风格总平均分布

驾驶风格按周期变化：驾驶风格相对同一个司机是固定的

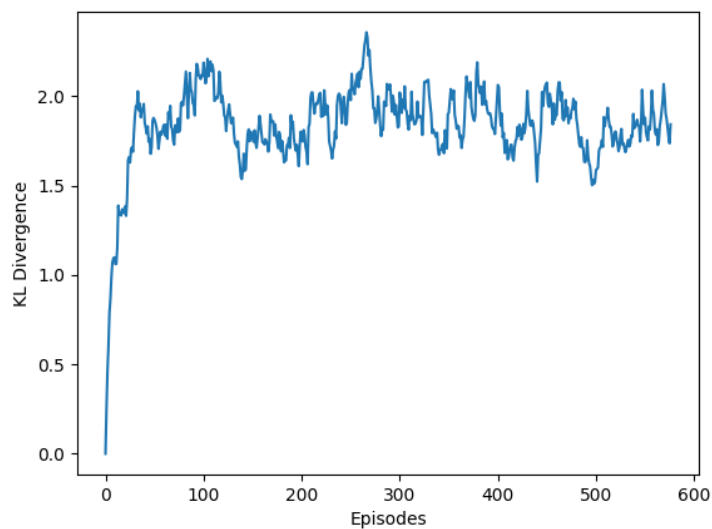


图 2 驾驶风格变化按 KL 散度评估, 风格相对固定

驾驶风格有 AI 和无 AI 比较

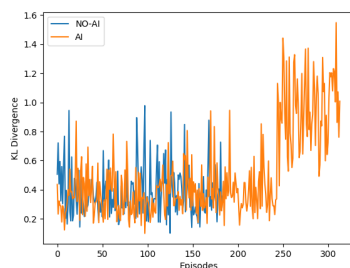


图 3.1 驾驶风格有 AI 和无 AI 比较, 后面打开 coastdown

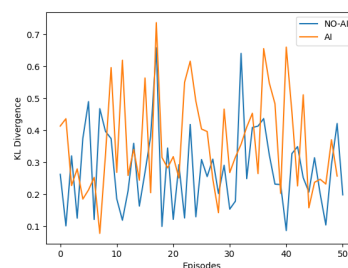


图 3.2 另一位驾驶员有 AI 与无 AI 比较

不同驾驶员风格以及统一驾驶员在应用不同算法后风格的定量比较

	SAC	DDPG-CD	SAC-CD	Driver 2-no CD
KLD	0	0.234	0.311	0.334

- 不同驾驶风格与 SAC 下驾驶风格总体比较:
  - KLD 可用于定量评估不同驾驶风格之间的差异
  - KLD 可用于监控训练过程中驾驶员风格和自己基准风格相比的变化

## 能耗

- 电动力默认 Pedal Map (PM) vs 自建 Pedal Map
  - 默认 PM: 高速时请求力矩会降低
  - 自建 PM: 分段线性, 请求力矩分段线性单调

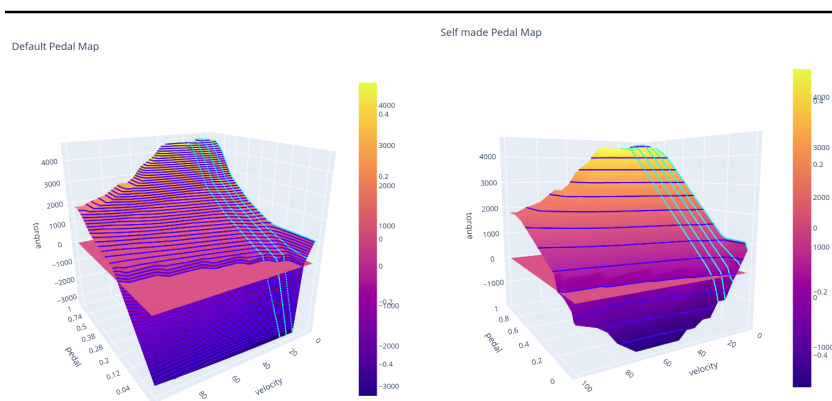


图 4.1 EP 默认 PM

图 4.2 自建 PM

- 默认 PM 和自建 PM 能耗比较
  - 自建 PM 作为初始表在每个训练开始时用于初始化
  - 自建 PM 对应的能耗作为比较的基准

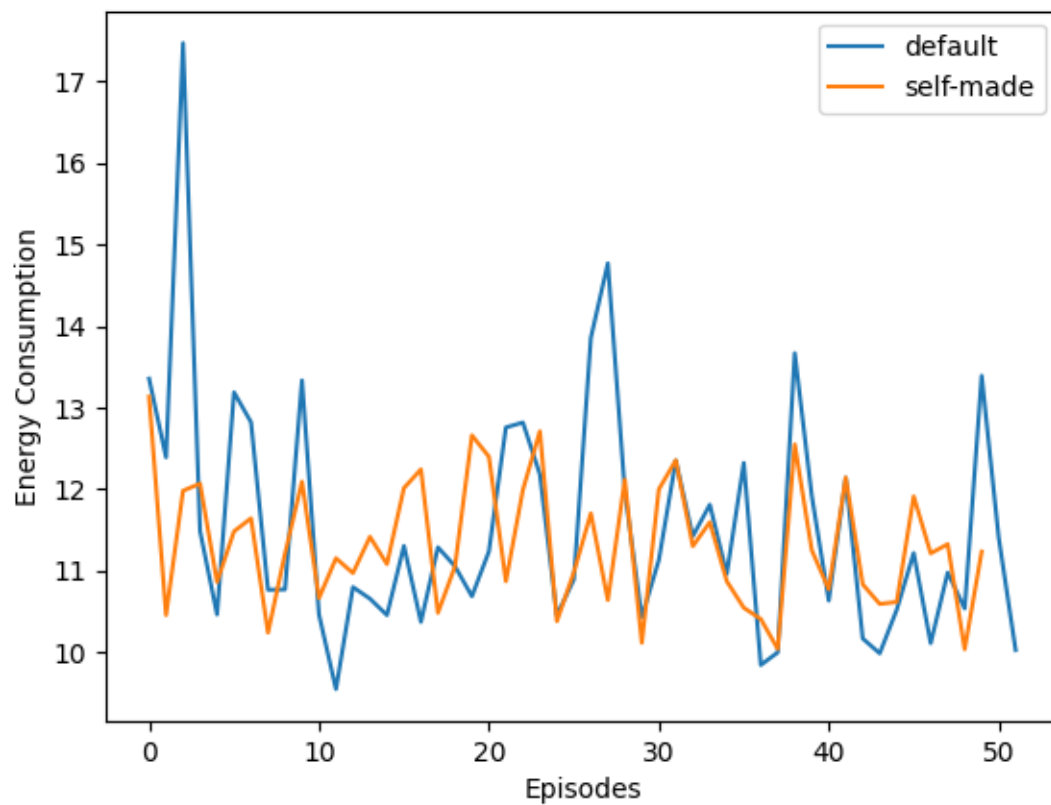


图 5 EP 默认 PM 与自建 PM 能耗比较,

## 能耗结果

历次带 AI tensorboard - 襄阳 vs. 上海 - 有时间同步问题, 成比例漏帧, 只影响测量, 大致不影响决策算法 - 确认收敛过程 - 能耗持续降低过程

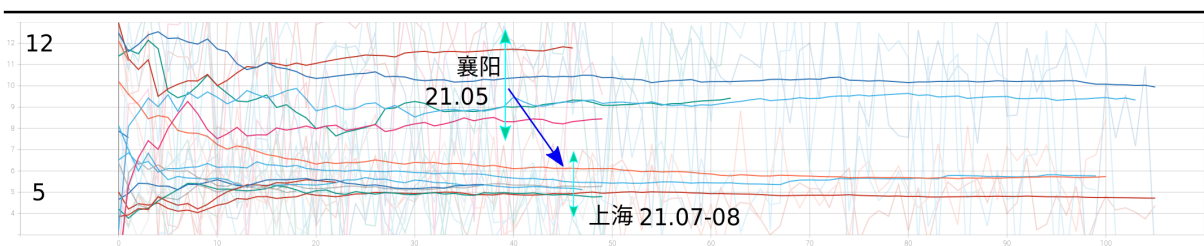


图 6 SAC 算法襄阳和上海对比

- 上海优化改进过程
  - 修复时间同步问题和漏帧问题
  - 能耗持续降低

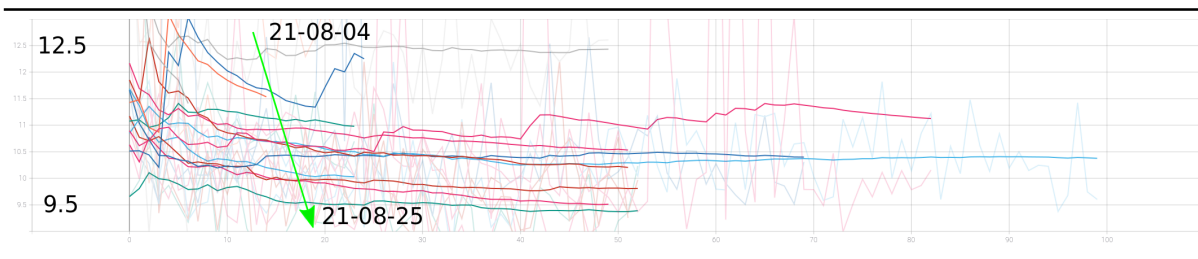


图 6 上海算法改进过程

- SAC(Stochastic Actor Critic) Pedal Map 非持续模式:

- 每个 epoch 使用上次 epoch 的模型,
- 开始 pedal map 用同一个默认表
- 模型继承之前的经验, 显示能耗持续降低

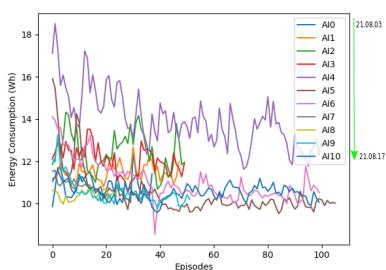


图 7.1 SAC 非持续模式能耗变化, 无 coastdown

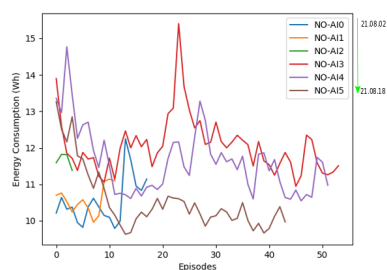


图 7.2 无 AI 模式能耗变化

- SAC Pedal Map 持续模式 (resume):

- 每个 epoch 使用之前的模型
- 开始 pedal map 用上一个 epoch 最后一个 episode 的表
- 模型继承之前经验, 且使用前一个训练周期的结果, 能耗结果趋近稳定

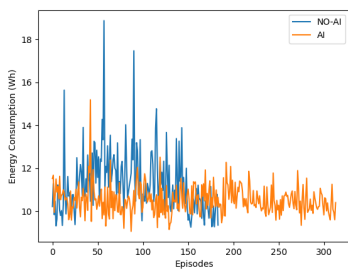


图 8.1 SAC 持续模式下能耗变化, 后面打开 coastdown, 原始数据

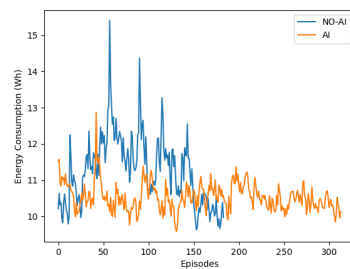


图 8.2 相同数据加平滑滤波

- SAC 对照组司机

- 在驾驶风格不变的情况下, 加入 SAC 算法使能耗降低
- 未打开 coastdown

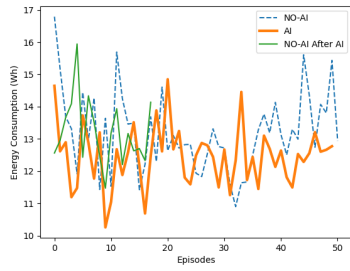


图 9.1 SAC 对照组能耗变化, 无 coastdown, 原始数据

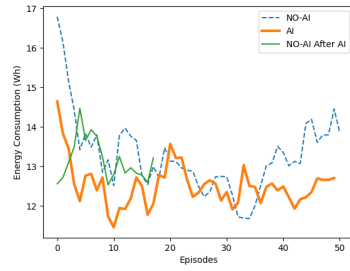


图 9.2 相同数据加平滑滤波

- SAC 偶发陷于局部最优
  - 未打开 Coast Down
  - 相当于随机策略收敛到一个确定性策略
  - 行动损失由于确定性策略下计算 logit 值, 趋向发散

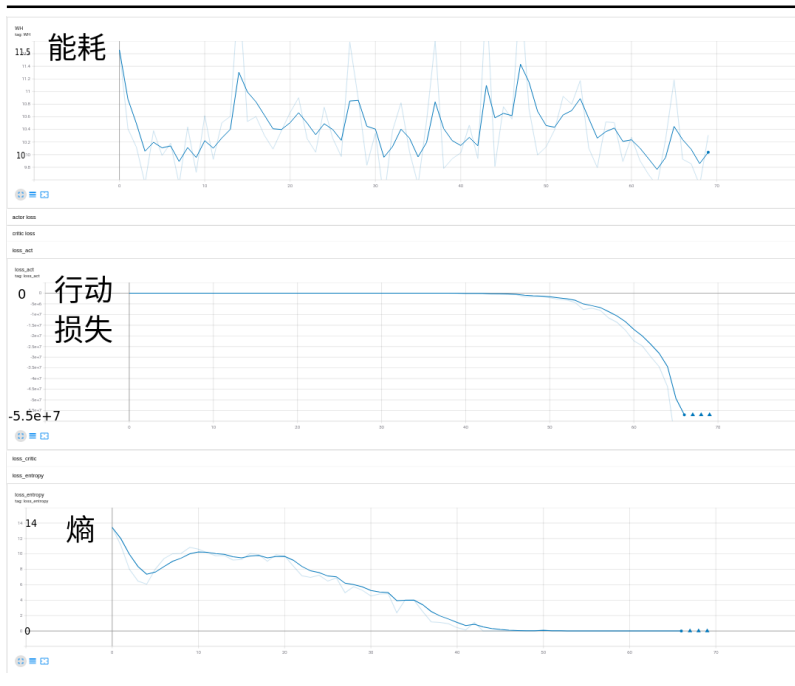


图 10 SAC 陷于确定性策略的局部最优, 随机策略的熵收敛到 0

- SAC 打开 Coastdown
  - 只打开 coastdown 动作空间, 并不使用专家知识有意利用 REGEN
  - 驾驶员和 agent 的合作决策
  - 收敛较快

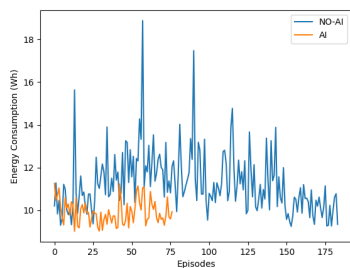


图 10.1 SAC 打开 coastdown, 原始数据

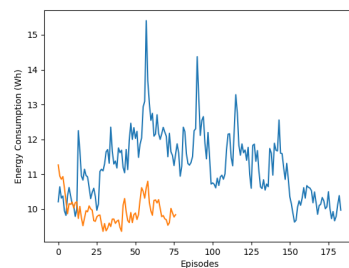


图 10.2 相同数据加平滑滤波

- DDPG 打开 Coast Down
  - 收敛更快, 大约是 SAC 的一倍
  - 同样的能耗改善结果,SAC 需要约 50 个 episode, DDPG 需要大约 25 个 episodes

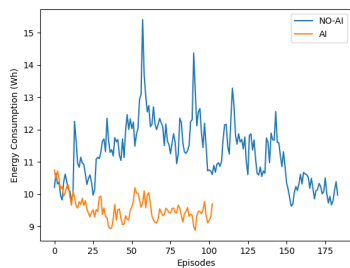


图 11.1 DDPG 打开 coastdown, 带平滑

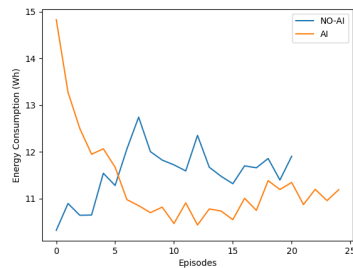


图 11.2 对照组司机数据, 带平滑

- DDPG Pedal Map 变化表
  - 随机采样策略现象
  - 对应能量回收的工况, 请求负扭矩变大

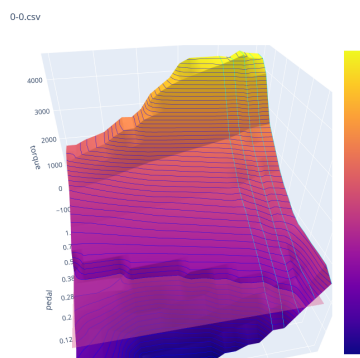


图 11.1 DDPG 打开 coastdown, 带平滑

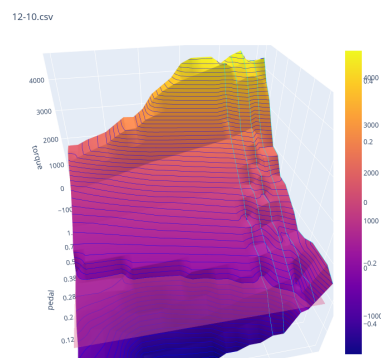


图 11.2 对照组司机数据, 带平滑

## 方法

强化学习方法, 以大数据为基础的奖励驱动优化方法

- **没有模型**
  - 车辆动力学的模型和知识
  - 电机模型
  - 电源管理系统模型
- 符合学习直觉:
  - 利用大数据建立内部模型
  - 自适应动态过程
- 下一步
  - 提高样本使用效率
    - \* 增加观测序列编码, 有助于利用更长时间序列的观测
    - \* 增加运动规划预测
  - 增加数据采集
    - \* 使用以大数据为基础的离线强化学习算法

\* 增加测试车辆