

基于强化学习的多无人机协同草原修复方法 研究 [4358754]

Xianyi Wang
Lanzhou University

Sunday 28th January, 2024

Contents

研究背景与意义

- ❑ 草原是地球上重要的生态系统，对生态和民生具有至关重要的作用。
- ❑ 中国草原面临退化和恶化趋势，需要保护和修复工作。
- ❑ 草原生态修复采用自然恢复和人工干预相结合的方式，根据地区情况进行规划。
- ❑ 无人机成为草原修复的重要工具，具有经济性和效率。
- ❑ 多无人机协同方法用于在能量有限的情况下最大化修复面积。

国内外研究现状

强化学习在路径规划中的应用：

- ❑ Liu等提出了将强化学习与遗传算法相结合，用于解决旅行商问题（TSP）。
- ❑ Bello等构建了指针网络进行强化学习，以应对TSP问题中网络输入规模的变化。
- ❑ Deudon等采用基于策略梯度的强化学习模型解决TSP问题。

多智能体强化学习：

- ❑ 孙长银等总结了多智能体强化学习方法的原理，并讨论了闭环框架和其中的重要问题及解决方法。
- ❑ 梁星星等探讨了多智能体强化学习中的关键技术，包括网络架构、模型鲁棒性、样本增强和对手建模等。

容量约束的车辆路径问题（CVRP）：

- ❑ Delarue等提出了基于价值学习的强化学习框架，用于解决CVRP问题。
- ❑ Lin等使用深度强化学习的路径规划应用方案，成功应用于受时间窗口限制的电动车辆路径规划任务。
- ❑ Li等考虑了不同车辆速度的差异，提出了基于深度强化学习的控制方案，解决了车辆异构的CVRP问题。
- ❑ Nazari等使用强化学习来训练输入受限的序列到序列结构网络，解决了站点需求随时间变化的CVRP问题。

研究内容

图 1: 主要研究内容

多无人机协同的草原修复问题

图 2: 多无人机修复退化区域实例

多无人机协同的草原修复面积最大化模型

序号	姓名	性别	年龄	身高/cm	体重/kg
1	张三	M	16	163	50
2	王红	F	15	159	47
3	李二	M	17	165	52

无人机能耗

无人机修复能耗:

$$e_i^S = \eta q_i, \quad (1)$$

$$q_i = (1 + l_i)^\gamma \quad (2)$$

$$E_u^S = \sum_{i=1}^N x_{iu} \omega_{iu} e_i^S \quad (3)$$

无人机信息收集能耗:

$$E_u^{ap} = e^{ap} \sum_{i=1}^N x_{iu} c_i \quad (4)$$

无人机飞行能耗:

$$E_u^f = e^f \sum_{i=0}^N \sum_{j=0}^N y_{iju} d_i^j, i \neq j, \quad (5)$$

$$e^f = \mathbb{M} \sqrt{\frac{g^3}{2\rho\zeta h}} \quad (6)$$

数学模型

该问题可被描述为:

$$\max_{x_{iu}, \sigma_{iu}} C^T = \sum_{u=1}^U \sum_{i=1}^N x_{iu} \sigma_{iu} \quad (7)$$

$$\text{s.t. } E_u^S + E_u^{ap} + E_u^f \leq E_{\max} \quad u \in U \quad (8)$$

$$\sum_{j=1}^N y_{0ju} = \sum_{i=1}^N y_{i0u} = 1, \quad u \in U \quad (9)$$

$$\sum_{j=1}^N y_{jiu} = \sum_{j=1}^N y_{iju} = x_{iu}, \quad \forall i \in M, u \in U, j \neq i \quad (10)$$

$$\sum_{u=1}^U x_{iu} = 1, \quad \forall i \in M \quad (11)$$

$$1 \leq \sigma_{iu} \leq c_i, \quad (12)$$

$$y_{iju} \in \{0, 1\}, \quad (13)$$

$$x_{iu} \in \{0, 1\} \quad (14)$$

问题建模

$$\text{状态} : t(< i), i \in [1, n+1] \quad (15)$$

$$\text{动作} : t(i) = (v_{i-1}, v_i) \quad (16)$$

$$\text{状态转移函数} : T(t(< i)|t(i)) = t(< i+1), i \in [1, n] \quad (17)$$

$$\text{随机策略函数} : p(t|V) = \prod_{i=1}^n p((t(i)|t(< i))) \quad (18)$$

$$\text{即时奖励} : r(t(< i)|t(i)) = d_{i-1}^i + \beta \frac{\sum_{j=1}^{\lceil i/2 \rceil} l_j}{\sum_{j=1}^i l_j} \quad (19)$$

$$\text{最终回报} : R(t|V) = d_n^0 + \sum_{i=0}^{n-1} \beta \frac{\sum_{j=1}^{\lceil i/2 \rceil} l_j}{\sum_{j=1}^i l_j} \quad (20)$$

网络结构

图 3: 网络结构图

网络目标

行动者网络:

$$J(\theta|V) = \mathbb{E}_{t \sim p_\theta(\cdot|V)} R(t|V) \quad (21)$$

$$\nabla_\theta J(\theta|V) = \mathbb{E}_{t \sim p_\theta(\cdot|V)} [(R(t|V) - b(V)) \nabla_\theta (t|V) \log_{p_\theta}(t|V)] \quad (22)$$

$$\nabla_\theta J(\theta) \approx \frac{1}{B} \sum_{i=1}^B (R(t|V_i) - b(V_i)) \nabla_\theta \log_{p_\theta}(t_i|V_i) \quad (23)$$

评论家网络:

$$L(\theta_c|V) = \frac{1}{B} \sum_{i=1}^B \|b_{\theta_c}(V_i) - R(t_i|V_i)\|_2^2 \quad (24)$$

协同调度算法

图 4: 多无人机修复退化区域实例

仿真设置

图 5: 仿真参数设置

仿真结果

图 6: 模拟训练过程

无人机修复详情

图 7: 无人机修复详情

无人机修复轨迹

图 8: 无人机修复轨迹

Thanks for your time!
Q&A