中国光谷·“华为杯”第十九届中国研究生数学建模竞赛

题 目 基于马尔可夫决策的车间生产调度问题

摘 要：

汽车制造业发展日新月异，越来越多的制造公司期待一种智能化的生产模式实现对车间汽车生产的调度控制。我们针对智能 PBS（Painted Body Store，汽车制造涂装-总装缓存调序区）调度策略问题，基于马尔可夫决策过程、时间差分方法、置信域策略优化以及蒙特卡洛方法等，针对于接车横移机与送车横移机的调度手段，建立与求解了 PBS 调度策略模型，并通过 Python 编程实现，能够很好的解决 PBS 的动态调度问题，并尽可能满足和提升生产需求以及系统效率。

针对问题一，建立 PBS 调度策略模型。首先，分别构建 PBS 状态空间编码，动作空间编码以及决策掩码，并量化优化目标与奖励变量。其中，决策掩码描述了接车横移机与送车横移机在不同情况下可以执行的动作，以满足题目中的约束条件。奖励变量描述了所执行动作对于优化目标产生的影响。然后，基于**马尔可夫决策过程**，建立 PBS 调度策略模型，并基于 Bellman 方程进行调度策略状态价值和横移车调度动作价值的迭代。为了降低求解的时间复杂度，我们采用了时间差分方法对 PBS 调度策略模型进行状态价值的求解。同时，我们摈弃了时间差分方法控制中的贪心采取调度策略的方法，引入了**置信域策略优化**来分解调度策略优化和调度状态价值评估，以保证了我们通过限制调度策略在迭代更新前后的 KL（Kullback-Leible）散度，从而能得到一个单调性能提升的调度策略。另外，考虑到算法的计算复杂度，我们进一步使用了裁剪技术来加快 PBS 调度策略的迭代更新。

最后，在模型评估中，第一，我们**对比了置信域策略优化和时间差分方法控制中贪心策略优化的差异**。实验结果表明，贪心策略会使得送车横移机容易陷入返回车道“陷阱”，即，其概率总在 0.5 左右波动。同时，贪心策略优化在这种情况下的 KL 散度也是非常局限，这说明了其很难探索到高效的 PBS 调度策略，而基于概率采样的置信域策略优化能更好的避免这种局部最优的情况，并使得 KL 散度是相对扩散的。第二，实验结果显示置信域策略优化下的 PBS 调度方案的**奖励是单调递增的**。第三，我们进一步**探索了优化目标权重设定影响**，由于题目中优化目标 2 的可实现频率远小于优化目标 1 和 3。因此，如果在题目的权重设定下，贪心策略易陷入局部最优。但对于置信域策略优化来说，它可以求解到最优的奖励权重以摆脱局部最优的问题，实验结果显示置信域策略优化求解到的优化目标的权重为 0.45，0.05，0.4 与 0.1 时，PBS 调度策略模型可以得到分数分别为 22.563（附录 1 数据）和 44.593（附录 2 数据）。

针对问题二，基于问题一建立的 PBS 调度策略模型，更改约束条件以验证模型效果。实验结果表明，PBS 调度策略分数为 23.733（附录 1 数据）和 46.428（附录 2 数据），同时在置信域策略优化下，PBS 调度方案的奖励仍然是单调递增的。除此之外，我们更改了实验的输入序列，模型仍能求解到有效的优化目标值，验证了所提模型具有鲁棒性与推广性。

关键词：PBS 调度策略模型；马尔可夫决策过程；时间差分方法；置信域策略优化