

DOI: 10.3901/JME.2014.21.084

基于部分可观测 Markov 决策过程理论的 盾构推进载荷规划*

胡祥涛¹ 张红旗¹ 李自成² 黄永安²

(1. 中国电子科技集团公司第 38 研究所 合肥 230088;

2. 华中科技大学机械科学与工程学院 武汉 430074)

摘要: 针对盾构掘进过程中位姿控制问题, 提出了基于部分可观测马尔科夫决策过程(Partially observable Markov decision processes, POMDP)理论的推进载荷规划方法。在推进载荷规划模型中, 将盾构自动纠偏看成不确定环境下序贯决策问题, 充分考虑掘进过程中随机因素的影响, 将盾构掘进过程中受到的阻力、推进载荷和盾构位姿分别定义为 POMDP 的状态集、行动集和观测集, 然后重点讨论了信念状态、状态转移函数和观测函数等几个关键参数的获取方法。在计算值函数时, 考虑了盾构位姿偏离程度和盾构载荷平稳程度对推进载荷决策的影响, 建立了立即收益函数和长期折扣收益函数, 并采用基于点的值迭代算法寻求推进载荷最优规划策略。针对天津地铁 9 号线进行了案例分析, 结果表明基于 POMDP 的推进载荷规划方法是合理有效的, 能够顺应掘进阻力随机变化。

关键词: 盾构法; 自动纠偏; POMDP; 推进载荷规划

中图分类号: TG156

POMDP-based Planning Model of Driving Force during Shield Tunneling

HU Xiangtao¹ ZHANG Hongqi¹ LI Zicheng² HUANG Yongan²

(1. No.38 Research Institute, China Electronics Technology Group Corporation, Hefei 230088;

2. School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074)

Abstract: In shield tunneling, the automatic control systems are mostly based on empirical relationships and do not have a precise theoretical background. Consequently, the present automatic control systems control the shield according to a snake-like motion to move it back onto the planned alignment. To solve such problems, an approach that adopts partially observable Markov decision processes (POMDP) theory to plan the driving force with explicit representation of the uncertainty during excavation is presented. In POMDP, the huge resistances during excavation, shield attitudes, and the driving force of possible world are scattered as the state set, observe set and action set, respectively. Furthermore, the deviations of shield attitude from planned alignment and the stationary of the loads acting on the shield are considered in the calculation of the value function. The experimental results confirm that the driving force planning based on POMDP is reasonable, and have certain ability of compliance at the mutation conditions.

Key words: shield tunneling; automatic deviation rectifying; POMDP; driving force planning

0 前言

为了实现极端环境下的高精、高效掘进作业,

盾构要求具备极高的轨迹跟踪能力, 如必须保证长达 20 km 的隧道轴线误差小于 20 mm。然而, 服役环境的随机多样性与施工引起的应力场、温度场、渗流场等多场耦合作用使得掘进界面的物理和力学行为极为复杂、难以预测, 导致盾构经常偏离设计轨迹, 严重情况下会造成重大工程事故^[1-4]。工程中,

* 国家重点基础研究发展计划(973 计划, 2013CB035405)和国家科技支撑计划 (2012BAF12B03)资助项目。20131109 收到初稿, 20140701 收到修改稿

工程师通常依赖施工经验或经验模型以类似蛇形的方式进行人工纠偏, 由于缺少理论背景, 盾构呈现蛇形运动趋势, 如图 1 所示, 影响了施工精度和效率^[4]。

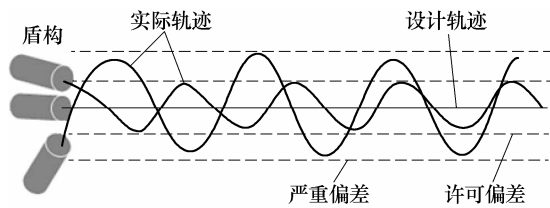


图 1 盾构蛇形运动轨迹示意图

近年来, 随着自动化技术的广泛应用, 国内外一些学者对盾构自动纠偏进行了广泛的研究, 如模糊纠偏控制系统、专家系统^[5-6]。文献[5]将卡尔曼滤波理论应用于盾构纠偏控制, 且利用自回归模型预测盾构动力学行为和千斤顶伸长量之间的关系, 从而建立反馈预测控制模型。在国内, 刘宣宇等^[7]系统地研究了盾构掘进系统及姿态自动控制技术, 全面总结了研究现状和未来发展趋势。李惠平等^[8]为了减少控制规则的数量并使控制器的性能易于调节, 提出一种“先分后合”的模糊控制器的设计方法, 仿真结果表明了方法的有效性。周奇才等^[9]引入模糊理论建立了盾构控制系统, 实现了盾构推进姿态的智能控制等功能, 极大地提高了施工效率和纠偏精度。

除了智能纠偏控制方法, 也有部分学者基于盾构动力学行为提出导向控制策略。文献[10-11]分析了盾构动力学行为, 并应用现代控制理论对盾构掘进机的控制进行了一系列研究, 得到盾构机的位置、姿态角与盾构纠偏载荷之间的关系。文献[12]研究了掘进过程中土-盾构相互作用, 将盾构看成欠驱动系统提出推进载荷控制策略。文献[13-14]研究了推进载荷力学模型, 提出了电液控制策略, 实现盾构掘进高效控制。

尽管目前众多学者对盾构自动纠偏问题做了大量有益探索工作, 但是由于施工过程中不确定因素的影响, 现有的自动纠偏技术仍然局限于试验与理论研究阶段, 难以实现工程应用。不同于以前的确定性研究方法, 本文充分考虑掘进过程中随机因素, 将盾构自动纠偏看成不确定环境下序贯决策问题, 采用部分可观测马尔科夫决策过程(Partially observable Markov decision processes, POMDP)理论研究推进载荷规划策略。

1 POMDP 概述

马尔科夫决策过程理论(Markov decision

processes, MDP)是一种研究不确定性环境中序贯决策问题的一门学科, 也是应用最广泛的决策模型^[15]。从 20 世纪 90 年代开始, 已经在人工智能、自动控制等领域中得到广泛的研究和应用。MDP 假设对于环境状态的观测是完全充分的, 即在每一步都确切地知道当前环境所处的状态。而通常系统无法做到这样的完全观测, 只能得到部分观测。基于对环境的部分观测的 Markov 决策过程称为 POMDP, 它是目前研究最多的决策论规划之一^[16-17]。

POMDP 的基本框架可以表示为 6 元数组 $\langle S, A, T, R, O, \Omega \rangle$ ^[18-20]。

(1) S 为有限状态集, 为智能体对环境及自身的状态信息。状态是智能体对自己所处环境的一种描述, 一般采用离散形式表示, 即对问题所有可能状态予以标号, 如 $S = \{1, 2, \dots, n\}$, n 代表了状态空间的规模。

(2) A 为有限行动集, 为智能体在任意时间节点上可以采取的行动, 类似于状态集, 通常表示为 $A = \{1, 2, \dots, m\}$, m 代表了行动空间的规模。

(3) T 为状态转换函数, $S \times A \rightarrow S'$ 表示在状态 s 下执行行动 a 后, 其后续状态为 s' 的可能概率分布 $\Pr(s'|s, a)$, 也记为 $T^a(s, s')$ 。

(4) R 为立即收益函数, $S \times A \rightarrow R$ 意为在状态 s 下执行行动 a 后期望得到的立即收益, 也称作报酬函数, 用 $R(s, a)$ 表示。

(5) O 为有限观测集, 为智能体可观测信息的集合, 是状态 S 的外在表现, 定义为 $O = \{1, 2, \dots, l\}$ 。

(6) Ω 为观测函数, $\Omega(s', a, o)$ 表示执行行动 a 后进入下一个状态 s' 时观测到 o 的概率, 即 $\Pr(o|s', a)$ 。

1.1 信念状态

POMDP 将环境的变迁看作一个状态空间 S 上的 Markov 链。对于智能体而言, 不可能知道环境当前的真实状态, 但是它可以收到一部分的观测信息 o 。由于观测是不完善的, 是部分信息, 因此智能体也不可能根据这个观测来完全确定当前的环境状态, 只能在一定程度上进行估计。这里的估计即称作信念状态。信念状态本质上是状态空间上的一个概率分布, 记为 $b(s)$, 有 $0 \leq b(s) \leq 1$, 且 $\sum_{s \in S} b(s) = 1$ 。

图 2 表示了 POMDP 决策过程。在这个模型里, 智能体首先计算当前的信念状态, 然后需要根据信念状态执行相应动作: $a = \pi(b)$, 并获取观测信息, 从而信念状态发生转移和实现整个过程的演变。

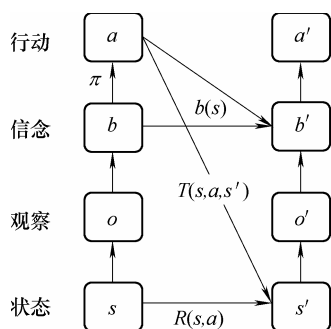


图 2 POMDP 模型中决策过程示意图

图 2 说明了信念状态之间的转移具有马尔可夫性。即在任意时刻 t , 智能体下一时刻的信念状态 b' 只与当前的信念状态 b 、行动 a 和观察 o' 有关, 而与之之前的历史无关。因此信念状态 b' 可以用贝叶斯公式进行更新。具体过程如下

$$b'(s') = \frac{\Pr(o'|s',a) \sum_{s \in S} \Pr(s'|a,s) \Pr(s|b)}{\Pr(o'|a,b)} = \frac{\Omega(s',a,o') \sum_{s \in S} T(s'|a,s)b(s)}{\Pr(o'|a,b)} \quad (1)$$

$$\Pr(o'|a,b) = \sum_{s' \in S} \Pr(o',s'|a,b) \Pr(s|b) = \sum_{s' \in S} \Pr(s'|a,b) \Pr(o'|s',a,b) = \sum_{s' \in S} \Omega(s',a,o') \sum_{s \in S} T(s',a,s)b(s) \quad (2)$$

式 (2) 为归一化因子。根据转移函数 T 及观测函数 Ω , 就可以实现信念状态的转移更新。

1.2 值函数

为了寻求最优策略, 需要确定值函数, 即目标函数。策略表示信念状态空间到行动集的映射: $a = \pi(b)$ 。值函数不是针对每个状态 s 进行计算的, 而是考虑在起始信念状态为 b 情况下的长期收益 (或总代价)。

给定一个策略, 就能够计算执行该策略所得到的长期收益, 但是长期收益值是智能体的初始状态有关的。对于一个初始状态 s , 按照策略执行动作得到的期望收益为

$$V(s) = R(s,a) + \gamma \cdot V_{future} \quad (3)$$

$$V_{future} = \sum_{s' \in S} T(s,a,s') \sum_{o' \in O} \Omega(s',o',a) V(s') \quad (4)$$

式中, V_{future} 表示执行动作 a 之后潜在长期收益; γ 为折扣值, 表示相对立即收益 R 存在一个时间折扣。

按照式 (3) 和式 (4), 初始信念状态 b 的最优值函数和最优策略可以表示为

$$V_t^*(b) = \max_{a \in A} [b(s)V(s)] = \max_{a \in A} \left[R(b,a) + \gamma \sum_{o' \in O} \Pr(o'|b,a) V_{t-1}^*(b') \right] \quad (5)$$

$$\pi_t^*(b) = \arg \max_{a \in A} V_t^*(b') \quad (6)$$

求解 POMDP 问题, 就是要计算整个信念状态空间中每个信念状态上的最优值函数——式 (5)。受“维度灾”的影响, 目前常采用的算法是基于点的值迭代算法 (Point-based value iteration, PBVI)。

PBVI 算法的每一次迭代过程由两个步骤组成: 第一步, 在给定的信念状态集合上更新值函数; 第二步, 扩充信念状态集合。这两个步骤交替执行, 直到值函数收敛或者满足终止条件为止^[19]。

2 推进载荷规划模型

盾构推进载荷规划是在一个极其复杂的环境下所做出的多阶段决策活动。为了顺应高达数十甚至上百倍变化的随机突变载荷, 必须在模型中引入掘进过程受到的掘进阻力信息, 使盾构能够在与环境的交互中不断增强推进载荷序贯决策能力, 保证盾构平稳掘进。显然, 掘进阻力是不能直接观测的状态信息, 属于隐藏状态。下面具体给出基于 POMDP 的盾构推荐载荷规划模型。

2.1 状态集、观测集与行动集

为描述盾构位姿, 首先建立整体坐标系: 竖直向下为 y 轴, 水平向前为 z 轴, x 轴由右手坐标系确定。从三维空间看, 盾构位姿可以用有向线段描述, 包含 6 个变量: 盾构位置定义为 (x,y,z) 、绕每个坐标轴转动的姿态角定义为 $(\beta, \alpha, \epsilon)$ 。一般情况下, 工程师主要关心的位姿是俯仰角 β 、偏转角 α 、和轴向位置 z , 相应的推进载荷包括推进力、俯仰力矩和偏转力矩。

盾构作业时, 操作工程师通过自动导向系统获得盾构当前运行的位姿状态。当工程师发现盾构位姿某个分量偏离设计轨迹时, 就会调整推进载荷中相应的分量。为简化分析, 本文假设位姿分量之间相互独立。接下来以俯仰角 β 为例介绍俯仰力矩的 POMDP 模型, 其他推进载荷分量的 POMDP 模型在相互独立假设条件下可以类似建立, 不再赘述。

(1) 状态集。状态是对盾构所处环境的整体描述。盾构掘进过程中所受到的阻力, 是不确定环境的直接反映, 因而本文将之定义为状态。

盾构掘进过程中, 监测系统记录了每个阶段的推进载荷。根据准静态运动关系, 可以通过对推进

载荷进行统计分析, 并考虑可能出现的突变载荷, 确定掘进阻力的波动范围。对应俯仰角, 将俯仰阻力矩在波动范围内离散, 定义为 POMDP 模型中的状态集: $S_\beta = \{m_1, \dots, m_n\}$ 。

(2) 行动集。为确定 POMDP 模型中的行动集, 需对实测推进载荷进行统计分析, 并考虑可能出现的突变载荷。对应俯仰角, 将 POMDP 模型中的行动集定义为离散的俯仰力矩: $A_\beta = \{M_1, \dots, M_m\}$ 。

(3) 观测集。受不确定因素影响, 俯仰角通常在一定范围内波动。将俯仰角离散, 即可定义为 POMDP 模型中的观测集: $O_\beta = \{\beta_1, \dots, \beta_l\}$ 。

2.2 转移函数和观测函数

转移函数和观测函数是 POMDP 模型中的两组重要参数, 本质上是两组条件概率。转移函数刻画了系统的状态演变规律, 而观测函数则描述了系统隐藏状态与观测信息之间的概率关系。对于实际问题, 转移函数和观测函数可以通过以下途径获得:

① 根据已发表的文献资料或者权威资料确定; ② 根据领域专家或者个人经验确定; ③ 由可获得的历史数据确定; ④ 根据模型随机模拟获得; ⑤ 综合使用上述方法确定^[14]。

本节中, 我们主要通过盾构历史掘进数据来确定转移函数和观测函数, 即通过同一线路已开挖隧道的掘进数据估计未开挖路段的统计特性。

(1) 转移函数。在人工纠偏过程中, 工程师根据测量获得的位姿偏离程度, 估计盾构受到的阻力, 然后做出推进载荷调整方案。当盾构在推进载荷作用下, 到下一个时间步时, 盾构受到的阻力要么还是跟以前一样, 要么将发生一定的变化。在 POMDP 模型中, 采用状态转移函数描述推进载荷调整前后状态的变化规律, 记作 $T(s_\beta, a_\beta, s'_\beta)$ 。根据前面介绍的 POMDP 模型的基本概念可知, 转移函数是一个条件概率, 即 $T(s_\beta, a_\beta, s'_\beta) = \Pr(s'_\beta | a_\beta, s_\beta)$, 其中 s_β 和 s'_β 分别表示前后时刻的状态, a_β 则表示当前推进载荷调整方案。

在盾构掘进过程中, 记录了实际推进载荷数据。因此, 可以根据历史推进载荷记录数据, 采用统计分析方法获得状态转移概率(在第 3 节, 将结合实例给出详细的统计分析过程, 这里不再赘述)。

(2) 观测函数。盾构受到的俯仰阻力矩是系统隐藏的真实状态, 而导向系统所监测到的盾构俯仰角是俯仰阻力矩和推进载荷(俯仰力矩)共同作用的结果。根据前文所述, POMDP 模型中的观测集为俯仰角, 则观测函数 $\Omega(a_\beta, s'_\beta, o_\beta)$ 表示采取行动 a_β

后, 盾构状态变为 s'_β 时, 测量系统可以得到观测 o_β 的条件概率, 即 $\Omega(a_\beta, s'_\beta, o_\beta) = \Pr(o_\beta | a_\beta, s'_\beta)$ 。

在盾构掘进过程中, 位姿测量系统会实时记录盾构俯仰角和推进载荷。根据盾构准静态掘进特征, 可统计出观测函数(在第 3 节, 将结合实例给出详细的统计分析过程, 这里不再赘述)。

2.3 值函数

立即收益函数是计算 POMDP 值函数的基本要素, 指当采取行动 a_β 后由一个状态 s_β 转变到另一个状态 s'_β 所得到的收益值(或代价值)。

在盾构纠偏过程中, 立即收益函数可以用纠偏效果的评价来定义。通常我们期望盾构位姿偏差越小越好, 同时期望盾构能够平稳运行, 避免载荷波动对设备造成损伤。因此, 立即收益函数应该包含盾构状态和观测信息。对于俯仰力矩 POMDP 模型, 立即收益函数定义如下

$$R(s_\beta, a_\beta) = E[K_e(o'_\beta - \beta^*)^2 + K_s(a_\beta - s'_\beta)^2] \quad (7)$$

式中 K_e ——盾构位姿偏离程度权重系数;

K_s ——盾构载荷平稳度权重系数;

β^* ——俯仰角设计值;

E ——数学期望。

式(7)第一项反映了盾构位姿偏离程度, 表明在决策载荷作用下, 俯仰角可能的偏离值; 第二项考虑了盾构受到的载荷平稳度, 表明了俯仰力矩和可能受到的阻力矩平衡程度。

POMDP 模型中的立即报酬依赖于盾构状态和观测信息, 因此我们可以通过实时监测数据计算每一个推进载荷的纠偏效果。例如, 某一个时刻盾构位姿偏差和载荷波动比较小, 可以认为盾构此时运行状态最佳; 相反位姿偏差很大, 载荷波动剧烈则认为盾构此时运行状态恶劣。通过权衡不同的运行状态, 设定 K_e 和 K_s 的取值。

将式(7)代入式(3)~(6), 即可得到俯仰力矩 PMDOP 模型的值函数和最优策略。注意, 在本文中值函数为最小化问题。

3 案例分析

为了评价 POMDP 模型规划推进载荷的能力, 我们将 POMDP 模型求出的规划策略与实际工程数据进行分析比较。试验所使用的数据是来自天津地铁 9 号线某标段, 该标段与水平线夹角约为 15.83

mrad), 即盾构向上俯仰掘进。每掘进一环, 记录了一组数据, 包括俯仰角和俯仰力矩。

3.1 状态集、观测集与行动集定义

图 3 和图 4 分别为天津地铁 9 号线某标段第 1 环~130 环对应的实测俯仰力矩和俯仰角, 图中横坐标皆为管片环数。

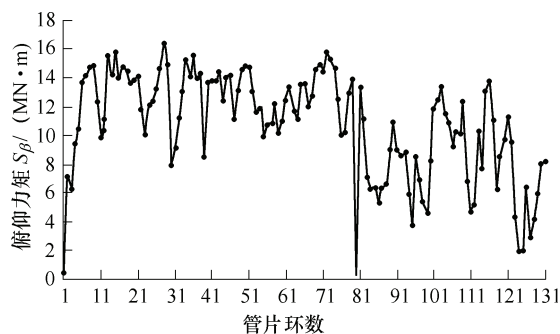


图 3 俯仰力矩实测数据

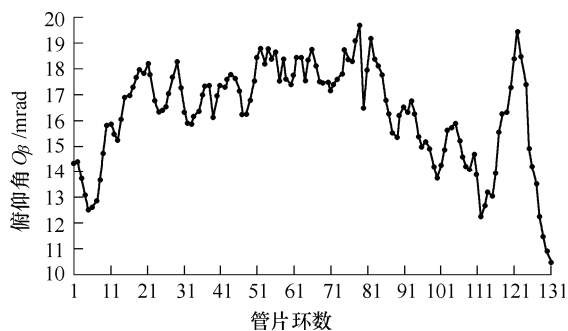


图 4 俯仰角实测数据

从图 3 和图 4 可以看出, 掘进过程中载荷和俯仰角波动比较大。由于准静态运动特征, 俯仰力矩和阻力矩的波动程度可近似认为是一致的。因此, 分别作下列定义: 状态集(即阻力矩) $S_\beta = \{6, 8, 10, 12, 14\}(\text{MN} \cdot \text{m})$, 决策集(即俯仰力矩) $A_\beta = S_\beta$, 将俯仰角定义为观测集 $O_\beta = \{14, 16, 18\}(\text{mrad})$ 。

3.2 转移函数和观测函数定义

在定义转移函数和观测函数之前, 需首先给出信念状态的定义方式。因为俯仰阻力矩是不能直接观测的, 只能根据 a_β 和 o_β 进行信念估计。本文根据施工经验, 定义信念状态如表 1~5 所示。以表 1 中数字 0.5 为例, 说明在决策值 $a_\beta = 6 \text{ MN} \cdot \text{m}$ 、观测值 $o_\beta = 14 \text{ mrad}$ 情况下, 对 $s_\beta = 8 \text{ MN} \cdot \text{m}$ 的估计概率为 0.5。这与工程实际相吻合, 当工程师观测到 $o_\beta = 14 \text{ mrad}$, 说明当前的平均阻力大于平均推进力, 故对 $s_\beta = 8 \text{ MN} \cdot \text{m}$ 的估计概率较高。注意, 盾构掘进是准静态过程, 存在某一瞬时阻力大于推进

力的情况。

根据表 1~5, 对图 3 所示俯仰力矩进行统计分析, 即可获取转移函数。统计分析方法如下: 如相邻时刻俯仰力矩分别为 $10 \text{ MN} \cdot \text{m}$ 和 $12 \text{ MN} \cdot \text{m}$, 对应的俯仰角分别为 14 mrad 和 16 mrad , 表明决策值 $a_\beta = 10 \text{ MN} \cdot \text{m}$, 当前阻力状态为 $b(s_\beta) = [0 \ 0 \ 0.1 \ 0.7 \ 0.2]$, 下一时刻阻力状态为 $b(s'_\beta) = [0 \ 0 \ 0.1 \ 0.8 \ 0.1]$, 则该次转移统计数值如表 6 所示。当统计完所有记录数据(第 1~131 环), 即可根据累计频次值获取状态转移函数, 如表 7~11 所示。

表 1 决策值 $a_\beta = 6 \text{ MN} \cdot \text{m}$ 时的信念状态

观测值 O_β/mrad	下一时刻状态值 $S_\beta/(\text{MN} \cdot \text{m})$				
	6	8	10	12	14
14	0.1	0.5	0.2	0.1	0.1
16	0.7	0.2	0.1	0	0
18	0.9	0.1	0	0	0

表 2 决策值 $a_\beta = 8 \text{ MN} \cdot \text{m}$ 时的信念状态

观测值 O_β/mrad	下一时刻状态值 $S_\beta/(\text{MN} \cdot \text{m})$				
	6	8	10	12	14
14	0	0.1	0.6	0.2	0.1
16	0.1	0.8	0.1	0	0
18	0.9	0.1	0	0	0

表 3 决策值 $a_\beta = 10 \text{ MN} \cdot \text{m}$ 时的信念状态

观测值 O_β/mrad	下一时刻状态值 $S_\beta/(\text{MN} \cdot \text{m})$				
	6	8	10	12	14
14	0	0	0.1	0.7	0.2
16	0	0.1	0.8	0.1	0
18	0.2	0.7	0.1	0	0

表 4 决策值 $a_\beta = 12 \text{ MN} \cdot \text{m}$ 时的信念状态

观测值 O_β/mrad	下一时刻状态值 $S_\beta/(\text{MN} \cdot \text{m})$				
	6	8	10	12	14
14	0	0	0	0.1	0.9
16	0	0	0.1	0.8	0.1
18	0.1	0.2	0.6	0.1	0

表 5 决策值 $a_\beta = 14 \text{ MN} \cdot \text{m}$ 时的信念状态

观测值 O_β/mrad	下一时刻状态值 $S_\beta/(\text{MN} \cdot \text{m})$				
	6	8	10	12	14
14	0	0	0	0.1	0.9
16	0	0	0.1	0.2	0.7
18	0.1	0.1	0.2	0.5	0.1

表 6 状态转移频次统计

当前状态 S_β / mrad	下一时刻状态 $S_\beta / (\text{MN} \cdot \text{m})$				
	6	8	10	12	14
6	0	0	0	0	0
8	0	0	0	0	0
10	0	0	0.01	0.08	0.01
12	0	0	0.07	0.56	0.07
14	0	0	0.02	0.16	0.02

表 7 决策值 $a_\beta = 6 \text{ MN} \cdot \text{m}$ 时的状态转移概率矩阵

当前状态 $S_\beta / (\text{MN} \cdot \text{m})$	下一时刻状态 $S_\beta / (\text{MN} \cdot \text{m})$				
	6	8	10	12	14
6	0.41	0.25	0.21	0.09	0.04
8	0.13	0.33	0.26	0.17	0.11
10	0.12	0.42	0.19	0.17	0.10
12	0.06	0.45	0.18	0.19	0.12
14	0.06	0.45	0.12	0.20	0.11

表 8 决策值 $a_\beta = 8 \text{ MN} \cdot \text{m}$ 时的状态转移概率矩阵

当前状态 $S_\beta / (\text{MN} \cdot \text{m})$	下一时刻状态 $S_\beta / (\text{MN} \cdot \text{m})$				
	6	8	10	12	14
6	0.40	0.19	0.33	0.05	0.03
8	0.21	0.45	0.12	0.06	0.16
10	0.07	0.35	0.12	0.09	0.37
12	0.04	0.32	0.12	0.10	0.42
14	0.04	0.32	0.12	0.10	0.42

表 9 决策值 $a_\beta = 10 \text{ MN} \cdot \text{m}$ 时的状态转移概率矩阵

当前状态 $S_\beta / (\text{MN} \cdot \text{m})$	下一时刻状态 $S_\beta / (\text{MN} \cdot \text{m})$				
	6	8	10	12	14
6	0.32	0.50	0.12	0.06	0
8	0.22	0.35	0.33	0.09	0.01
10	0.04	0.18	0.22	0.45	0.11
12	0	0.12	0.10	0.37	0.41
14	0	0.12	0.08	0.34	0.46

表 10 决策值 $a_\beta = 12 \text{ MN} \cdot \text{m}$ 时的状态转移概率矩阵

当前状态 $S_\beta / (\text{MN} \cdot \text{m})$	下一时刻状态 $S_\beta / (\text{MN} \cdot \text{m})$				
	6	8	10	12	14
6	0.20	0.33	0.20	0.23	0.04
8	0.17	0.29	0.31	0.20	0.03
10	0.16	0.26	0.30	0.21	0.07
12	0.09	0.09	0.26	0.27	0.29
14	0.04	0.14	0.14	0.17	0.51

表 11 决策值 $a_\beta = 14 \text{ MN} \cdot \text{m}$ 时的状态转移概率矩阵

当前状态 $S_\beta / (\text{MN} \cdot \text{m})$	下一时刻状态 $S_\beta / (\text{MN} \cdot \text{m})$				
	6	8	10	12	14
6	0.16	0.17	0.18	0.41	0.08
8	0.14	0.15	0.28	0.36	0.07
10	0.13	0.13	0.25	0.37	0.12
12	0.12	0.13	0.25	0.37	0.13
14	0.05	0.06	0.16	0.31	0.42

观测函数可以通过类似统计方法确定。如某时刻俯仰力矩为 $10 \text{ MN} \cdot \text{m}$ 、俯仰角为 14 mrad ，表明在 $a_\beta = 10 \text{ MN} \cdot \text{m}$ 作用下，观测到 $o_\beta = 14 \text{ mrad}$ ，则该次观测按 $b(s_\beta) = [0 \ 0 \ 0.1 \ 0.7 \ 0.2]$ 进行离散统计。当统计完所有记录数据(第 1~131 环)，即可根据累计频次值获取观测函数，如表 12~16 所示。

表 12 决策值 $a_\beta = 6 \text{ MN} \cdot \text{m}$ 时的观测概率矩阵

当前状态 $S_\beta / (\text{MN} \cdot \text{m})$	俯仰角观测值 O_β / mrad		
	14	16	18
6	0.13	0.50	0.37
8	0.78	0.17	0.05
10	0.79	0.21	0
12	1	0	0
14	1	0	0

表 13 决策值 $a_\beta = 8 \text{ MN} \cdot \text{m}$ 时的观测概率矩阵

当前状态 $S_\beta / (\text{MN} \cdot \text{m})$	俯仰角观测值 O_β / mrad		
	14	16	18
6	0	0.22	0.78
8	0.13	0.83	0.04
10	0.88	0.12	0
12	1	0	0
14	1	0	0

表 14 决策值 $a_\beta = 10 \text{ MN} \cdot \text{m}$ 时的观测概率矩阵

当前状态 $S_\beta / (\text{MN} \cdot \text{m})$	俯仰角观测值 O_β / mrad		
	14	16	18
6	0	0.10	0.90
8	0.08	0.79	0.13
10	0.85	0.15	0
12	1	0	0
14	0	0	1

表 15 决策值 $a_\beta = 12 \text{ MN} \cdot \text{m}$ 时的观测概率矩阵

当前状态 $S_\beta / (\text{MN} \cdot \text{m})$	俯仰角观测值 O_β / mrad		
	14	16	18
6	0	0	1
8	0	0	1
10	0	0.09	0.91
12	0.03	0.80	0.17
14	0.75	0.25	0

表 16 决策值 $a_\beta = 14 \text{ MN} \cdot \text{m}$ 时的观测概率矩阵

当前状态 $S_\beta / (\text{MN} \cdot \text{m})$	俯仰角观测值 O_β / mrad		
	14	16	18
6	0	0	1
8	0	0	1
10	0	0.17	0.83
12	0.03	0.13	0.84
14	0.32	0.50	0.18

3.3 值函数计算

根据图 3 和图 4, 在掘进到第 11~20 环时, 俯仰力矩和俯仰角相对比较平稳, 以该段数据预估式 (7) 中权重系数 K_e 和 K_s 。预估方法如下: 计算俯仰角与设计值的平均偏差, 及俯仰力矩与阻力矩的平均偏差; 在不考虑单位的情况下, 前者约为后者的 1/3。故令式 $K_e=1$ 、 $K_s=1/3$, 表明姿态平稳性和载荷平稳性同等重要。

将 $K_e=1$ 和 $K_s=1/3$ 代入式 (7), 计算立即收益 $R(s_\beta, a_\beta)$ 得

$$R(s_\beta, a_\beta=6)=[6.00 \quad 8.96 \quad 8.60 \quad 9.28 \quad 9.28]$$

$$R(s_\beta, a_\beta=8)=[4.44 \quad 4.82 \quad 7.68 \quad 8.23 \quad 8.23]$$

$$R(s_\beta, a_\beta=10)=[6.36 \quad 4.92 \quad 4.44 \quad 5.87 \quad 6.16]$$

$$R(s_\beta, a_\beta=12)=[8.15 \quad 7.81 \quad 7.53 \quad 5.26 \quad 5.01]$$

$$R(s_\beta, a_\beta=14)=[11.05 \quad 10.85 \quad 10.02 \quad 9.93 \quad 6.21]$$

将上述计算结果代入式 (5), 在 Matlab 里采用 PBVI 算法求解, 即可获取最优策略。

3.4 结果分析

图 5 给出了基于 POMDP 规划策略俯仰力矩规划结果和实测结果, 其中实测结果来源于天津地铁 9 号线某标段 131~148 环对应的数据。从图中可以看出俯仰力矩规划值与实测值变化趋势相吻合, 表明了基于 POMDP 的推进载荷规划方法是合理有效的, 能够顺应掘进阻力随机变化。

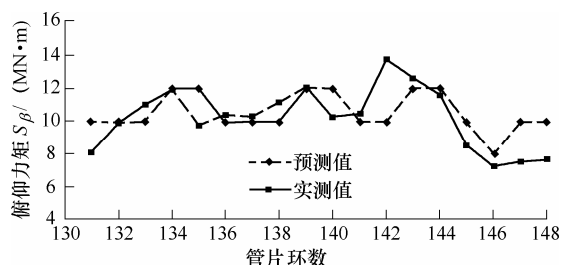


图 5 基于 POMDP 俯仰力矩规划结果

此外, 规划结果与实测结果相比——最大偏差发生在 142 环, 约 $3\text{MN} \cdot \text{m}$, 相对误差约在 25% 以内。误差来源主要有三个方面。首先, 信念状态定义误差: 在 POMDP 模型中, 一般基于工程经验定义信念状态, 但是经验具有很大的不确定性, 将直接导致结果误差。其次, 集合定义误差: 在 POMDP 模型中, 状态集、观测集和行动集采用的是离散形式, 但本文采用的集合规模偏小, 离散间隔取值偏大, 影响了载荷规划精度。最后, 数据统计误差: 工程中, 每掘进一环只记录了一组数据, 导致样本量偏小, 影响了转移函数和观测函数的有效性。因此, 为进一步提高规划精度, 不仅需要足够多的掘进数据, 还需要给出更合理的信念状态定义方法,

以及大规模 POMDP 问题的有效求解算法。

4 结论

本文针对盾构自动纠偏问题, 提出了基于 POMDP 理论的推进载荷规划方法。当自动纠偏被看成不确定环境下序列决策问题时, POMDP 模型可以帮助发现最优推进载荷规划策略。在构建推进载荷规划模型时, 重点讨论了如何获得状态转移函数、观测函数和立即收益函数等 POMDP 建模过程中的几个关键步骤。最后, 针对天津地铁 9 号线进行了案例分析, 结果表明基于 POMDP 的推进载荷规划方法是合理有效的, 能够顺应掘进阻力随机变化。

然而, 本章建立的推进载荷规划模型时仅仅考虑了一种理想的变化情况, 对掘进过程很多问题做了一些必要的简化。虽然模型具有一定的顺应性和有效性, 但是该模型仍然存在很多需要进一步改进的地方, 比如细化状态的表示、在模型中加入其他决策评价方法、更合理的转移函数和观测函数获取方法、针对行动空间和观察空间庞大的问题提出更高效的 POMDP 近似算法等。

参考文献

- [1] HU Xiangtao, HUANG Yongan, YIN Zhouping, et al. Optimization-based model of tunneling-induced distributed loads acting on the shield periphery[J]. Automation in Construction, 2012, 24(1): 138-148.
- [2] 胡国良, 龚国芳, 杨华勇. 基于压力流量复合控制的盾构推进液压系统[J]. 机械工程学报, 2006, 42(6): 124-127.
- [3] HU Guoliang, GONG Guofang, YANG Huayong. Thrust hydraulic system of shield tunnel boring machine with pressure and flow compound control[J]. Chinese Journal of Mechanical Engineering, 2006, 42(6): 124-127.
- [4] 邓颖聪, 郭为忠, 高峰. 盾构推进系统分区性能分析的等效机构建模[J]. 机械工程学报, 2010, 46(3): 122-127.
- [5] DENG Yingcong, GUO Weizhong, GAO Feng. Equivalent mechanism-based modeling for grouping performance analysis of the thrust system of shield machines[J]. Chinese Journal of Mechanical Engineering, 2010, 46(3): 122-127.
- [6] HU Xiangtao, HUANG Yongan, YIN Zhouping, et al. Driving force planning in shield tunneling based on Markov decision processes[J]. Sci. China Tech. Sci., 2012, 55(4): 1022-1030.

- [5] SAKAI K, HOSHUTANI M. Prediction and control of behaviors on driving shields using Kalman filter theory[J]. J. Japan. Soc. Civ. Eng. , 1987, 385(VI-7): 69-78.
- [6] KUWAHARA H, HARADA M. Study of applicability for fuzzy control theory in shield tunneling process[C]// Proceedings of Civil Society, Japan, 1988: 169-178.
- [7] 刘宣宇, 邵诚. 盾构机自动控制技术现状与展望[J]. 机械工程学报, 2010, 46(20): 152-160.
LIU Xuanyu, SHAO Cheng. Present status and prospect of shield machine automatic control technology[J]. Journal of Mechanical Engineering, 2010, 46(20): 152-160.
- [8] 李惠平, 夏明耀. 盾构姿态的模糊控制方法[J]. 同济大学学报, 2003, 31(7): 824-827.
LI Huiping, XIA Mingyao. Fuzzy algorithm for control of shield machine[J]. Journal of Tongji University, 2003, 31(7): 824-827.
- [9] 周奇才, 陈俊儒, 何自强, 等. 盾构智能化姿态控制器的设计[J]. 同济大学学报, 2008, 36(1): 76-80.
ZHOU Qicai, CHEN Junru, HE Ziqiang, et al. Design of intellectualized controller of shield machine[J]. Journal of Tongji University, 2008, 36(1): 76-80.
- [10] SUGIMOTO M, SRAMOON A. Theoretical model of shield behavior during excavation I: Theory[J]. Journal of Geotechnical and Geoenvironmental Engineering, 2002, 128(2): 138-155.
- [11] SRAMOON A, SUGIMOTO M, KAYUKAWA K. Theoretical model of shield behavior during excavation II : Application[J]. Journal of Geotechnical and Geoenvironmental Engineering, 2002, 128(2): 156-165.
- [12] HUANG Yongan, YIN Zhoupin, HU Xiangtao, et al. A novel force control method for quasistatic underactuated multibody systems[J]. Advanced Robotics, 2009, 23(10): 1249-1260.
- [13] SHI Hu, YANG Huayong, GONG Guofang, et al. Determination of the cutterhead torque for EPB shield tunneling machine[J]. Automation in Construction, 2011, 1(1): 27-29.
- [14] YANG Huayong, SHI Hu, GONG Guofang, et al. Electro-hydraulic proportional control of thrust system for shield tunneling machine[J]. Automation in Construction, 2009, 18(7): 950-956.
- [15] PUTERMAN M L. Markov decision processes: Discrete stochastic dynamic programming [M]. New York: John Wiley & Sons Inc., 1994.
- [16] MONAHAN G E. A survey of partially observable Markov decision processes : Theory , models , and algorithms[J]. Management Science, 1982, 28(1): 1-16.
- [17] WILLIAMS J D, YOUNG S. Partially observable Markov decision processes for spoken dialog systems[J]. Computer Speech & Language, 2007, 21(2): 393-422.
- [18] 范长杰. 基于马尔可夫决策理论的规划问题的研究[D]. 合肥: 中国科学技术大学, 2008.
FAN Changjie. Research on planning based on Markov decision theory[D]. Hefei: University of Science and Technology of China, 2008.
- [19] 冯奇. POMDP 近似解法研究及在中医诊疗方案优化中的应用[D]. 北京: 北京交通大学, 2011.
FENG Qi. Research on approximate algorithms of POMDP and application to TCM therapy planning[D]. Beijing: Beijing Jiaotong University, 2011.
- [20] ITOH H, NAKAMURA K. Partially observable Markov decision processes with imprecise parameters[J]. Artificial Intelligence, 2007, 171(8-9): 453-490.

作者简介: 胡祥涛, 男, 1981 年出生, 博士。主要研究方向为载荷规划、可靠性分析、数字化设计。

E-mail: hust_hoo@163.com

张红旗, 男, 1975 年出生, 硕士, 研究员。主要研究方向为数字化设计与制造。

E-mail: vp_38@163.com

李自成, 男, 1977 年出生, 博士。主要研究方向为智能控制、自动导向。

E-mail: lizich@sohu.com

黄永安(通信作者), 男, 1981 年出生, 博士, 副教授。主要研究方向为柔性电子制造、多体动力学与控制。

E-mail: yahuang@hust.edu.cn