

基于启发式动态规划的盾构土压平衡优化控制

刘宣宇^{*1}, 许胜¹, 张凯举¹, 曹雨濛²

(1. 辽宁石油化工大学 信息与控制工程学院, 辽宁 抚顺 113001;

2. 吉林大学 数学学院, 吉林 长春 130012)

摘要: 为了避免盾构施工造成地表变形而引发安全事故,建立有效的密封舱土压平衡自动控制方法是至关重要的.为此,提出了一种新的基于启发式动态规划(heuristic dynamic programming, HDP)的盾构土压平衡优化控制方法,该方法能够实现盾构机多参数实时协调优化控制.基于神经网络,构建组成密封舱土压 HDP 控制器的执行网络、模型网络和评价网络,采用梯度下降算法对网络权值进行在线训练;利用评价网络迭代逼近密封舱土压优化控制代价函数,将多级优化问题简化为单级优化,从而显著降低实时计算负荷;执行网络以最小化代价函数为目标,对多参数同步协调优化,获得推进速度、刀盘扭矩、总推力、螺旋输送机转速的最优值.仿真结果表明,该方法能够有效控制密封舱土压平衡, HDP 控制器具有较强的抗干扰能力,并且控制过程更加快速、稳定.

关键词: 土压平衡盾构;密封舱土压;启发式动态规划;神经网络

中图分类号: TP23;U455.43 **文献标识码:** A **doi:** 10.7511/dllgxb201805013

0 引言

随着城市化进程不断加快,土压平衡盾构已被广泛应用于软土地层的地下工程建设.各地地质条件不同,对盾构施工过程中盾构密封舱土压控制要求也就不同.施工过程中,盾构密封舱土压无法得到有效控制,易导致地表变形和严重的安全事故.因此,实现土压平衡盾构的密封舱土压平衡控制,是避免地表变形和保障施工安全的关键所在.

土压平衡盾构主要通过调整螺旋输送机转速,改变密封舱渣土体积,维持密封舱土压与开挖面压力平衡,有效控制地表变形^[1].王林涛等^[2]提出基于前馈-密封舱压力反馈的土压控制方法,通过调节推进速度实现土压平衡.张晓峰^[3]以螺旋输送机转速为控制量,利用神经网络算法优化控制量,实现土压平衡控制.曹丽娟等^[4]提出基于PID神经元的密封舱土压控制方法. Yang 等^[5]利用传统PID控制器控制密封舱土压,取得了较好控制效果.但以上控制方法均是以单变量优化控

制密封舱土压,没有考虑其他控制参数的影响作用,密封舱土压控制精度有待进一步提高.

启发式动态规划(heuristic dynamic programming, HDP)是自适应动态规划的一种,是解决传统动态规划“维数灾”问题的一种近似最优控制方法^[6].其利用函数近似结构逼近代价函数,通过离线迭代或在线更新方式获得系统的最优代价函数及最优控制律,能有效解决非线性系统的多变量优化控制问题.其已被应用于倒立摆平衡控制^[7]、水泥立磨生料细度控制^[8]、储能系统优化控制^[9]等实际系统中,成为现代科学理论与工程领域中对复杂系统研究的一种重要方法.

因此,本文提出基于HDP的密封舱土压优化控制方法.通过对盾构掘进经验数据分析,发现刀盘扭矩是影响密封舱土压平衡的重要因素,两者之间存在非线性依赖关系^[10-11].为提高密封舱土压预测精度,本文将刀盘扭矩作为模型输入之一,建立密封舱土压预测模型,进而构建密封舱土压优化控制代价函数和HDP控制器.最后,通过实验仿真验证所提出方法的有效性.

1 HDP 基本结构与原理

密封舱土压 HDP 控制器结构如图 1 所示。

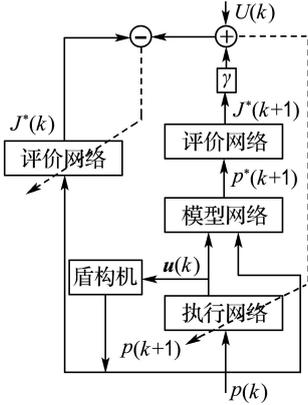


图 1 密封舱土压 HDP 控制器结构

Fig.1 Structure of HDP controller for soil pressure in sealed cabin

图 1 中, $p(k)$ 为 k 时刻密封舱土压; $p(k+1)$ 为 $k+1$ 时刻密封舱土压; $\gamma \in (0, 1]$, 为折扣因子; $\mathbf{u}(k) = (n(k) \ v(k) \ F(k) \ T(k))^T$, 为 k 时刻控制向量, 其中 $F(k)$ 、 $n(k)$ 、 $v(k)$ 、 $T(k)$ 分别为总推力、螺旋输送机转速、推进速度、刀盘扭矩; $U(k) = f(p(k), \mathbf{u}(k), k)$, 为密封舱效用函数, 对密封舱土压和控制向量做出评价; $J(\cdot)$ 为密封舱土压优化控制代价函数。

各信号流向如实线所示, 评价网络与执行网络的误差反馈路径如虚线所示。密封舱土压作为执行网络的输入, 输出为控制向量; 模型网络的输入为密封舱土压和控制向量, 输出为下一时刻密封舱土压, 其作为评价网络的输入, 输出密封舱土压优化控制代价函数近似值。评价网络实现密封舱土压优化控制代价函数的近似, 执行网络以最小化密封舱土压优化控制代价函数为目标, 优化控制向量。各网络详细训练流程和参数优化过程, 见下文 HDP 控制器设计部分。

HDP 的基本原理是采用迭代形式训练评价网络以近似 Bellman 动态规划方程中的代价函数:

$$J(k) = \sum_{i=0}^{\infty} \gamma^i U(k+i) \quad (1)$$

其中 $J(k)$ 表示效用函数从 k 时刻开始的迭代值。通过折扣因子进行折算, 显示效用函数迭代的时间效应。HDP 的目标是选择最优控制向量, 使得式(1)取极小值。

盾构机为高度复杂的非线性系统, 本文假设盾构系统为下式所示的离散系统:

$$p(k+1) = f(p(k), \mathbf{u}(k), k) \quad (2)$$

对于盾构系统式(2), 式(1)可表示为

$$J(k) = \sum_{i=k}^{\infty} \gamma^{i-k} U(p(i), \mathbf{u}(i), i) \quad (3)$$

式(3)经过推导可表示为

$$J(k) = U(k) + \gamma J(k+1) \quad (4)$$

根据 Bellman 优化原理, 若以式(4)作为网络权值调整的目标, 则可以通过网络训练使得代价函数逐步趋向当前控制策略下的最优值。本文以式(4)作为密封舱土压优化控制的代价函数, 设计基于 HDP 的土压平衡控制器。为了使评价网络近似 $J(k)$, 定义如下评价网络误差:

$$e_c(k) = J^*(k) - \gamma J^*(k+1) - U(k) \quad (5)$$

$$E_c(k) = \frac{1}{2} e_c^2(k) \quad (6)$$

其中 $J^*(k)$ 为 k 时刻代价函数估计值, $J^*(k+1)$ 为 $k+1$ 时刻代价函数估计值。

执行网络的训练以最小化式(4)为目标, 获得最优控制向量。因此, 定义如下执行网络误差:

$$e_a(k) = \gamma J^*(k+1) + U(k) \quad (7)$$

$$E_a(k) = \frac{1}{2} e_a^2(k) \quad (8)$$

2 HDP 控制器设计

2.1 密封舱土压预测模型建立

王洪新等^[12-13]通过对盾构掘进过程机理分析, 推导出盾构总推力、密封舱土压、螺旋输送机转速、推进速度等掘进参数间的关系表达式。根据该研究结果, 设定本文的密封舱土压预测模型的输入变量为当前密封舱土压 $p(k)$ 、当前螺旋输送机转速 $n(k)$ 、当前推进速度 $v(k)$ 、当前总推力 $F(k)$ 、当前刀盘扭矩 $T(k)$; 输出为下一时刻密封舱土压 $p(k+1)$ 。即有

$$p(k+1) = f(p(k), n(k), v(k), F(k), T(k)) \quad (9)$$

BP 神经网络因其强大的非线性映射能力, 被广泛用于复杂工业控制系统的建模。本文将双极性函数和线性函数作为隐藏层和输出层的传递函数; 隐藏层节点个数选择对模型的训练效果至关重要, 因此本文首先统计了隐藏层节点个数为 5~16 的土压预测模型的土压预测误差; 通过 Matlab 仿真验证, 当网络隐藏层节点个数为 9 时, 密封舱

土压预测误差为 10.5%，而其他的均超过 12.2%。因而，构建基于 3 层 BP 神经网络的密封舱土压预测模型，结构如图 2 所示。

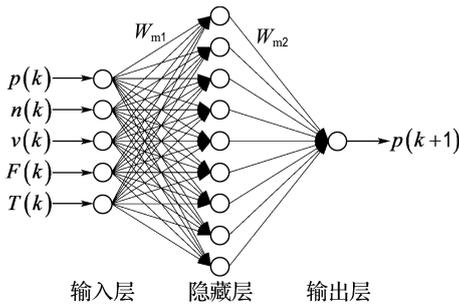


图 2 密封舱土压预测模型

Fig. 2 Forecasting model of soil pressure in sealed cabin

2.2 密封舱土压效用函数

密封舱土压效用函数是整个 HDP 控制器设计过程中与密封舱土压直接相关的重要指标。实际上，在自适应动态规划中，设计效用函数的过程本质上就是优化设计 HDP 控制器的过程。在某种程度上，效用函数的选择决定了控制器动态控制的优劣。在对工业系统的控制过程中，效用函数必须能够反映实际系统的控制特效和系统自身属性，所设计的控制器才可满足控制要求。本文控制对象是密封舱土压，目的是使其稳定于 0.18~0.22 MPa，因此选择的效用函数为

$$U(k) = \begin{cases} [p(k) - 0.2]^2; & |p(k) - 0.2| \leq 0.02 \\ 1; & \text{其他} \end{cases} \quad (10)$$

2.3 评价网络的训练

隐藏层节点个数的选择方法同 2.1 节，通过仿真验证，当隐藏层节点个数为 8 时，代价函数收敛值为 0.068，其他节点个数的代价函数收敛值大于 0.1。因此，本文评价网络采用 1-8-1 结构的 3 层 BP 神经网络。评价网络的训练流程如图 3 所示，各信号流向如实线所示，反向传播路径如虚线所示。

根据 Bellman 优化原理， k 时刻密封舱土压控制器的最优代价函数可表示为

$$J^*(k) = \min_{u(k)} \{U(p(k), u(k), k) + \gamma J^*(k+1)\} \quad (11)$$

由式(5)、(6)定义的误差，根据梯度下降算法和链式法则，评价网络隐藏层到输出层的权值更新规则如下：

$$\Delta W_{e2}(k) = l_c \left[-\frac{\partial E_c(k)}{\partial W_{e2}(k)} \right] = l_c \left[-\frac{\partial E_c(k)}{\partial J^*(k)} \frac{\partial J^*(k)}{\partial W_{e2}(k)} \right] \quad (12)$$

$$W_{e2}(k+1) = W_{e2}(k) + \Delta W_{e2}(k) \quad (13)$$

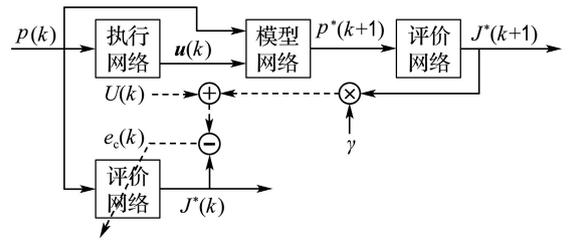


图 3 评价网络的训练流程

Fig. 3 Training process of critic network

输入层到隐藏层的权值更新规则如下：

$$\Delta W_{e1}(k) = l_c \left[-\frac{\partial E_c(k)}{\partial W_{e1}(k)} \right] \quad (14)$$

$$W_{e1}(k+1) = W_{e1}(k) + \Delta W_{e1}(k) \quad (15)$$

式中： $l_c \in (0, 1]$ ，为评价网络学习率； $\Delta W_c(k)$ 为 k 时刻评价网络权值增量； $W_{e1}(k)$ 为 k 时刻评价网络输入层到隐藏层的权值； $W_{e2}(k)$ 为 k 时刻评价网络隐藏层到输出层的权值； $W_c(k+1)$ 为 $k+1$ 时刻评价网络权值。

2.4 模型网络的训练

本文模型网络采用如图 2 所示的 3 层 BP 神经网络结构。将该模型网络输出与上文基于 BP 神经网络的密封舱土压预测模型式(9)输出之间产生的误差 $e_m(k+1)$ ，用于模型网络权值更新。

根据梯度下降算法和链式法则，通过最小化下式定义的模型网络误差实现权值更新：

$$e_m(k+1) = p(k+1) - p^*(k+1) \quad (16)$$

其中 $p(k+1)$ 为密封舱土压预测模型输出， $p^*(k+1)$ 为模型网络输出。

$$E_m(k+1) = \frac{1}{2} e_m^2(k+1) \quad (17)$$

同理，模型网络隐藏层到输出层的权值更新规则如下：

$$\Delta W_{m2}(k) = l_m \left[-\frac{\partial E_m(k+1)}{\partial W_{m2}(k)} \right] \quad (18)$$

$$W_{m2}(k+1) = W_{m2}(k) + \Delta W_{m2}(k) \quad (19)$$

输入层到隐藏层的权值更新规则如下：

$$\Delta W_{m1}(k) = l_m \left[-\frac{\partial E_m(k+1)}{\partial W_{m1}(k)} \right] \quad (20)$$

$$W_{m1}(k+1) = W_{m1}(k) + \Delta W_{m1}(k) \quad (21)$$

式中： $l_m \in (0, 1]$ ，为模型网络学习率； $\Delta W_m(k)$ 为

k 时刻模型网络权值增量; $W_{m1}(k)$ 为 k 时刻模型网络输入层到隐藏层的权值; $W_{m2}(k)$ 为 k 时刻模型网络隐藏层到输出层的权值; $W_m(k+1)$ 为 $k+1$ 时刻模型网络权值.

2.5 执行网络的训练

隐藏层节点个数的选择方法同 2.1 节, 通过仿真验证, 相较于其他节点个数, 当隐藏层节点个数为 8 时, 由执行网络优化获得的控制向量更加近似实际参数值. 因此, 本文执行网络采用 1-8-4 结构的 3 层 BP 神经网络. 执行网络目标是最小化评价网络输出的密封舱土压优化控制代价函数, 获得最优控制向量. 执行网络利用反向传播的评价网络误差, 实现其权值更新. 执行网络的训练流程如图 4 所示, 各信号流向如实线所示, 反向传播路径如虚线所示.

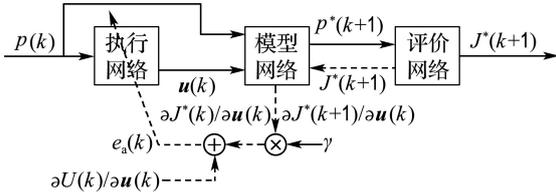


图 4 执行网络的训练流程

Fig. 4 Training process of action network

根据 Bellman 优化原理, k 时刻的最优土压控制向量可表示为

$$u^*(k) = \arg \min_{u(k)} \{U(p(k), u(k), k) + \gamma J^*(k+1)\} \quad (22)$$

根据式(7)、(8)定义的误差, 同理, 执行网络隐藏层到输出层的权值更新规则如下:

$$\Delta W_{a2}(k) = l_a \left[-\frac{\partial E_a(k)}{\partial W_{a2}(k)} \right] = l_a \left[-\frac{\partial E_a(k)}{\partial J^*(k)} \frac{\partial J^*(k)}{\partial u(k)} \frac{\partial u(k)}{\partial W_{a2}(k)} \right] \quad (23)$$

$$W_{a2}(k+1) = W_{a2}(k) + \Delta W_{a2}(k) \quad (24)$$

其中 $\frac{\partial J^*(k)}{\partial u(k)} = \frac{\partial U(k)}{\partial u(k)} + \gamma \frac{\partial J^*(k+1)}{\partial u(k)}$.

输入层到隐藏层的权值更新规则如下:

$$\Delta W_{a1}(k) = l_a \left[-\frac{\partial E_a(k)}{\partial W_{a1}(k)} \right] \quad (25)$$

$$W_{a1}(k+1) = W_{a1}(k) + \Delta W_{a1}(k) \quad (26)$$

式中: $l_a \in (0, 1]$, 为执行网络学习率; $\Delta W_a(k)$ 为 k 时刻执行网络权值增量; $W_{a1}(k)$ 为 k 时刻执行网络输入层到隐藏层的权值; $W_{a2}(k)$ 为 k 时刻执行网络隐藏层到输出层的权值; $W_a(k+1)$ 为 $k+1$

时刻执行网络权值.

2.6 HDP 控制器训练策略

(1) 初始各网络的权值, 赋值各网络学习率和折扣因子, 并设置最大迭代步数 q_{max} .

(2) 设定初始密封舱土压 $p(k)$. 将其作为执行网络输入, 输出控制向量 $u(k) = (n(k) \quad v(k) \quad F(k) \quad T(k))^T$.

(3) 将 $p(k)$ 输入评价网络, 输出密封舱土压优化控制代价函数 $J^*(k)$; 同时, 将当前密封舱土压 $p(k)$ 与控制向量 $u(k)$ 输入模型网络, 获得 $k+1$ 时刻密封舱土压 $p^*(k+1)$.

(4) 求解效用函数 $U(k)$.

(5) 将 $p^*(k+1)$ 作为评价网络输入, 输出下一时刻密封舱土压优化控制代价函数 $J^*(k+1)$.

(6) 依据评价网络误差 $e_c(k)$, 更新网络权值, 逼近密封舱土压优化控制代价函数.

(7) 执行网络根据评价反馈的误差 $e_a(k)$, 实现权值更新, 优化控制向量.

(8) 判断迭代误差. 在误差允许范围内, 执行下一步; 反之, 返回(5), 循环更新网络权值.

(9) 判断迭代步数, 若 $q \leq q_{max}$, 更新状态和控制向量, 返回(3); 否则, 训练结束.

3 仿真实验

3.1 预测模型训练及仿真

为建立密封舱土压预测模型, 需要足够多的样本数据才能反映实际掘进过程. 本文实验所用数据来源于北京某地铁施工现场的实测数据. 密封舱土压数据通过 4 个安置于密封舱隔板的压力传感器获取, 传感器分布如图 5 所示. 本文以传感器 1 的压力数据为例进行仿真实验.

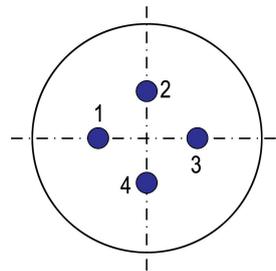


图 5 密封舱隔板压力传感器分布

Fig. 5 Pressure sensor distribution of the seal bulkhead

本次实验仿真采集了 800 组实测数据, 经过数据处理, 剔除一部分异常数据后得到 650 组有效数据, 用其中的 550 组作为训练样本, 另外 100

组作为测试样本. 通过 Matlab 对密封舱土压预测模型进行仿真. 有无刀盘扭矩的密封舱土压预测模型仿真如图 6 所示; 密封舱土压预测误差如图 7 所示; 两种预测模型均方误差如表 1 所示.

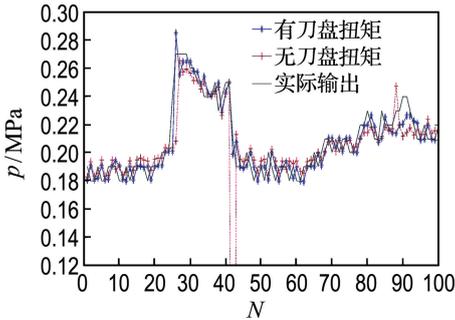


图 6 有无刀盘扭矩的密封舱土压预测模型仿真
Fig. 6 Simulation of soil pressure prediction model for sealed cabin with or without cutterhead torque

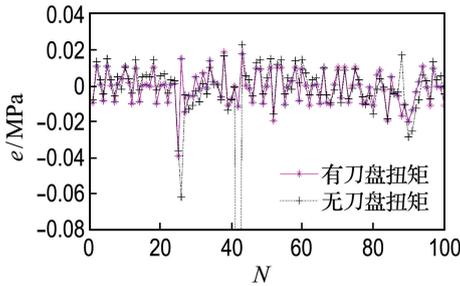


图 7 密封舱土压预测误差

Fig. 7 Prediction error of soil pressure in sealed cabin

表 1 预测模型均方误差

Tab. 1 Mean square error of prediction models

密封舱土压预测模型	均方误差
有刀盘扭矩	9.1554×10^{-5}
无刀盘扭矩	3.2000×10^{-3}

由表 1 可知, 当刀盘扭矩作为模型控制量之一时, 其均方误差更小, 说明考虑刀盘扭矩的土压预测模型具有更好的预测精度. 如图 6、7 所示, 考虑刀盘扭矩的模型, 其预测值与实测值整体拟合效果较好. 由于地质条件或工况的不确定性, 个别点拟合效果较差, 但误差基本在 ± 0.02 MPa (10%) 以内, 满足工程要求.

3.2 HDP 控制器仿真

根据本文 HDP 控制器设计部分, 通过 Matlab 对提出的控制方法进行实验仿真. 系统初始密封舱土压 $p(k) = 0.18$ MPa; l_a 、 l_c 初始值为 0.7; 折扣因子 $\gamma = 1$.

密封舱土压优化控制代价函数轨迹如图 8 所示; 密封舱土压优化轨迹如图 9 所示; 相应的, 刀盘扭矩、总推力、推进速度、螺旋输送机转速的优化轨迹如图 10 所示.

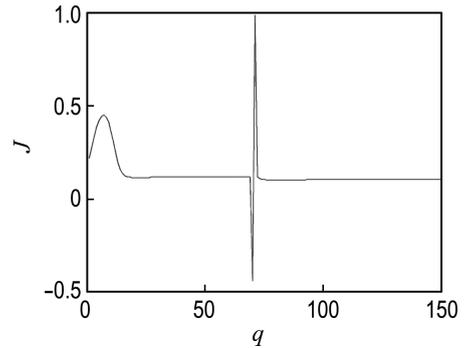


图 8 密封舱土压优化控制代价函数轨迹

Fig. 8 Cost function trajectory of optimization control of soil pressure in sealed cabin

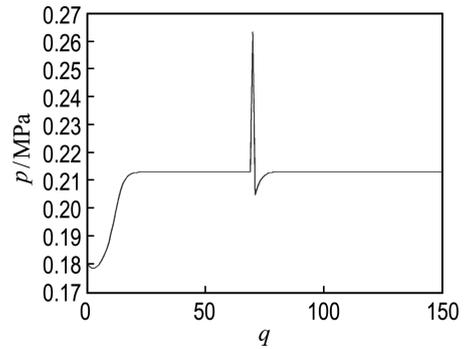


图 9 密封舱土压优化轨迹

Fig. 9 Optimization trajectory of soil pressure in sealed cabin

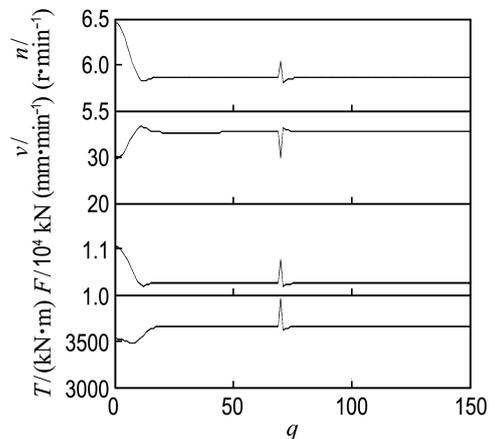


图 10 控制量优化轨迹

Fig. 10 Optimization trajectory of control variables

为了验证 HDP 控制器对密封舱土压优化效果, 本文采用文献[3]中盾构土压平衡控制器进行

仿真对比,其密封舱土压控制轨迹如图 11 所示。

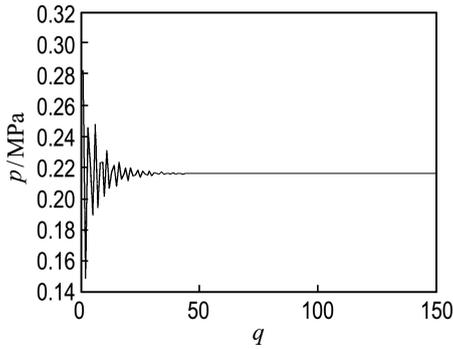


图 11 盾构土压平衡控制器土压控制轨迹

Fig. 11 Soil pressure control trajectory of controller of soil pressure balance of shield

通过对土压优化和控制轨迹图 9、11 对比,可以看出,相较于文献[3]中设计的控制器,HDP 控制器对密封舱土压优化调节时间更短,超调更小,优化控制过程更加稳定。文献[3]盾构土压平衡控制器仅可对螺旋输送机转速单变量控制,而 HDP 控制器可实现对刀盘扭矩、总推力、推进速度、螺旋输送机转速多变量同步协调控制。对 4 个控制量的调整过程及相应密封舱土压变化趋势,符合实际盾构施工时控制参数变化情况。

为了验证 HDP 控制器抗干扰能力,在迭代 70 步时加入干扰。如图 8 所示,密封舱土压优化控制代价函数仍能快速收敛,说明 HDP 控制器对密封舱土压优化控制是有效的,具有较好的动态性能。如图 9、10 所示,加入干扰后密封舱土压很快达到稳定状态,4 个控制量同样达到最优,说明 HDP 控制器具有较强的抗干扰能力。

4 结 语

本文基于 BP 神经网络将刀盘扭矩作为预测模型控制量之一,建立密封舱土压预测模型。与不考虑刀盘扭矩模型相比,考虑刀盘扭矩能够有效提高预测模型的预测精度。针对密封舱土压难以稳定控制的问题,在上述密封舱土压预测模型的基础上,本文提出基于 HDP 的密封舱土压优化控制方法。相较于单变量控制,该方法可实现对刀盘扭矩、总推力、推进速度、螺旋输送机转速的多变量控制;并且该方法收敛速度更快,稳定性较好,具有强抗干扰能力。基于 HDP 的密封舱土压控制器,能够满足密封舱土压的控制要求,达到稳定土压目的,为实现盾构掘进过程的优化控制提

供了一种新途径。如何提高密封舱土压模型的预测精度并增强 HDP 控制器的鲁棒性及其跟踪效果是今后的研究重点。

参 考 文 献:

- [1] LI Shouju, YU Shen, QU Fuzheng. Modeling approaches to pressure balance dynamic system in shield tunneling [J]. *Journal of Central South University*, 2014, **21**(3):1206-1216.
- [2] 王林涛,龚国芳,杨华勇,等. 基于前馈-反馈复合控制的盾构土压平衡控制[J]. *中南大学学报(自然科学版)*, 2013, **44**(7):2726-2735.
WANG Lintao, GONG Guofang, YANG Huayong, et al. Earth pressure balance control based on feedforward-feedback compound control [J]. *Journal of Central South University (Science and Technology)*, 2013, **44**(7):2726-2735. (in Chinese)
- [3] 张晓峰. 基于神经网络的盾构土压平衡控制模型的研究[D]. 大连:大连理工大学, 2009.
ZHANG Xiaofeng. Control model of earth pressure for shield tunneling based on neural network [D]. Dalian: Dalian University of Technology, 2009. (in Chinese)
- [4] 曹丽娟,上官子昌,李守巨,等. 基于神经网络的盾构机密封舱压力控制研究[J]. *广西师范大学学报:自然科学版*, 2010, **28**(3):175-181.
CAO Lijuan, SHANGGUAN Zichang, LI Shouju, et al. Pressure control for headchamber of shield machine based on neural network [J]. *Journal of Guangxi Normal University: Natural Science Edition*, 2010, **28**(3):175-181. (in Chinese)
- [5] YANG Huayong, SHI Hu, GONG Guofang, et al. Earth pressure balance control for EPB shield [J]. *Science in China Series E-Technological Sciences*, 2009, **52**(10):2840-2848.
- [6] 张化光,张欣,罗艳红,等. 自适应动态规划综述[J]. *自动化学报*, 2013, **39**(4):303-311.
ZHANG Huaguang, ZHANG Xin, LUO Yanhong, et al. An overview of research on adaptive dynamic programming [J]. *Acta Automatica Sinica*, 2013, **39**(4):303-311. (in Chinese)
- [7] MALLA N, NI Z. A new history experience replay design for model-free adaptive dynamic programming [J]. *Neurocomputing*, 2017, **266**: 141-149.
- [8] 林小峰,孔伟凯. 基于 ELM 的水泥立磨生料细度 ADP 控制[J]. *系统仿真学报*, 2016, **28**(11):2764-2770.

- LIN Xiaofeng, KONG Weikai. Adaptive dynamic programming in raw meal fineness control of vertical mill grinding process based on extreme learning machine [J]. *Journal of System Simulation*, 2016, **28**(11):2764-2770. (in Chinese)
- [9] 李相俊,张晶琼,何宇婷,等. 基于自适应动态规划的储能系统优化控制方法[J]. 电网技术, 2016, **40**(5):1355-1362.
- LI Xiangjun, ZHANG Jingqiong, HE Yuting, *et al.* Optimal control method of energy storage system based on adaptive dynamic programming [J]. *Power System Technology*, 2016, **40**(5):1355-1362. (in Chinese)
- [10] SHI Hu, YANG Huayong, GONG Guofang, *et al.* Determination of the cutterhead torque for EPB shield tunneling machine [J]. *Automation in Construction*, 2011, **20**(8):1087-1095.
- [11] 曹曦,邵诚,安毅. 盾构密封舱土压预测及智能优化控制策略[J]. 煤炭学报, 2015, **40**(12):2979-2986.
- CAO Xi, SHAO Cheng, AN Yi. Earth pressure prediction and optimal control intelligent strategy for shield chamber [J]. *Journal of China Coal Society*, 2015, **40**(12):2979-2986. (in Chinese)
- [12] 王洪新,傅德明. 土压平衡盾构掘进的数学物理模型及各参数间关系研究[J]. 土木工程学报, 2006, **39**(9):86-90.
- WANG Hongxin, FU Deming. A mathematical model and the related parameters for EPB shield tunneling [J]. *China Civil Engineering Journal*, 2006, **39**(9):86-90. (in Chinese)
- [13] 刘宣宇,邵诚,刘任喜. 盾构掘进土压控制模型的建立及仿真[J]. 煤炭学报, 2010, **35**(4):575-579.
- LIU Xuanyu, SHAO Cheng, LIU Renxi. Modeling and simulation on earth pressure control for shield tunneling [J]. *Journal of China Coal Society*, 2010, **35**(4):575-579. (in Chinese)

Optimization control for soil pressure balance of shield based on heuristic dynamic programming

LIU Xuanyu^{*1}, XU Sheng¹, ZHANG Kaiju¹, CAO Yumeng²

(1. School of Information and Control Engineering, Liaoning Shihua University, Fushun 113001, China;
2. School of Mathematics, Jilin University, Changchun 130012, China)

Abstract: In order to avoid the safety accidents caused by surface deformation in shield construction, it is very important to establish an effective automatic control method for the soil pressure balance of sealed cabin. For that, a new optimization control method for soil pressure balance of shield is proposed based on heuristic dynamic programming (HDP), which can realize coordinated optimization control of shield machine's multi-parameter in real time. Firstly, based on neural network, the action network, model network and critic network of the HDP controller for soil pressure in sealed cabin are established, and the gradient descent algorithm is used to train the network weights online. Then, the critic network is used to iteratively approximate the cost function of soil pressure control in sealed cabin, so that the multi-stage optimization problem is simplified to single-stage optimization, which can greatly reduce the calculation load. Finally, with the goal of minimizing the cost function of soil pressure control, the action network realizes the synchronous coordinated optimization of multi-variable such as advancing speed, cutterhead torque, total thrust and screw conveyor speed. The simulation results show that the method can effectively control the soil pressure balance of sealed cabin, and the HDP controller has strong anti-interference ability. What's more, the control process is quicker and steadier.

Key words: earth pressure balance shield; soil pressure in sealed cabin; heuristic dynamic programming; neural network