DOI: 10.7641/CTA.2013.21083

启发式动态规划在污水处理过程控制中的应用

薄迎春[1,2†], 乔俊飞[2]

(1. 中国石油大学(华东) 信息与控制工程学院, 山东 青岛 266580; 2. 北京工业大学 电子信息与控制工程学院, 北京 100124)

摘要: 针对溶解氧及硝态氮浓度的跟踪控制问题, 提出了一种基于回声状态网络的启发式动态规划控制方法, 该方法首先对当前策略进行评价, 然后根据评价结果对当前策略进行调整, 这个过程交替进行, 直至发现最优的控制策略. 评价函数及控制策略的逼近均采用回声状态网络实现. 为保证控制器的可用性, 对控制器学习过程的参数选择范围进行了分析. 污水处理过程的控制实验表明, 该方法能够显著提高系统控制的平稳性及控制精度.

关键词: 启发式动态规划; 污水处理; 溶解氧; 回声状态网络; 收敛性

中图分类号: TP273　　**文献标识码**: A

Application of heuristic dynamic programming to wastewater treatment process control

BO Ying-chun[1,2†], QIAO Jun-fei[1]

(1. College of Information and Control Engineering, China University of Petroleum (East China), Qingdao Shandong 266580, China; 2. College of Electronic and Control Engineering, Beijing University of Techonology, Beijing 100124, China)

Abstract: To control the dissolved oxygen concentration and nitrate concentration of wasterwater treatment process (WWTP), we propose a heuristic dynamic programming (HDP) method based on the echo state network (ESN). This scheme evaluates the current policy, and then the HDP controller adjusts the current policy according to the evaluation results. This procedure is continued until the optimal policy is found. Echo state networks are adopted to approximate the evaluation function and the optimal policy. To ensure the availability of the HDP controller, we investigate the parameter ranges of the HDP controller. Experimental results indicate that the ESN-based HDP controller has advantages over the PID controller in control stability and control precision.

Key words: heuristic dynamical programming (HDP); wastewater treatment; dissolved oxygen; echo state network (ESN); convergence

1　引言(Introduction)

污水处理是关系环境可持续发展的关键问题之一. 污水处理过程非线性较强, 各控制目标间耦合严重, 入水流量及污染物浓度频繁变化, 这些因素增加了污水处理过程控制的复杂性, 也导致常规控制效果较差[1]. 近年来, 模型预测控制(model predictive control, MPC)在该领域得到了一定的研究[2–3]. 如文献[2]基于MPC方法对污水处理过程的溶解氧浓度进行控制. 为提高系统的除氮性能, 其在MPC的基础上增加了入水氨氮的前馈补偿, 该方法取得了较好的除氮效果. 但是, 由于污水处理过程模型的时变特性和反应过程的不确定性, 建立适合控制需求的污水处理过程模型较为困难[3], 这也增加了MPC在污水处理领域应用的难度[4].

启发式动态规划(heuristic dynamic programming, HDP)[5–6]是解决模型难以建立的控制问题的有效方法之一. HDP控制器的设计过程对系统的动力学模型依赖较小, 甚至可以在无模型的方式下进行[7]. 典型的HDP控制器主要由评价模块和控制模块两大部分组成[5]. 首先给定一个控制策略, 系统在该策略下发生状态转移, 评价模块根据状态的反馈信息对当前的控制策略的效果进行评价(策略评价), 然后控制模块根据评价结果对当前的策略进行调整(策略提升), 这个过程不断重复, 直至寻找到最优的控制策略. 由于HDP能够在与环境的交互中逐渐学习到最优的策略, 所以HDP本质上属于一种自适应的学习控制方法. 目前, HDP已经在发电[8]、复杂非线性系统控制[9]、交通[10]等多个领域得到了研究和应用.

针对污水处理过程模型难以建立的问题, 提出了一种基于回声状态网络的HDP控制方法, 并将其应用

收稿日期: 2012−10−20; 收修改稿日期: 2013−02−17.

†通信作者. E-mail: boyingchun@sina.com.cn; Tel.: +86 18811315642.

基金资助: 国家自然科学基金重点资助项目(61034008); 北京市自然科学基金资助项目(4092010).

于污水处理过程的溶解氧和硝态氮浓度控制.

2 污水处理过程简介(Review of wastewater treatment process)

工业和生活污水经过一系列前置处理措施进入污水处理装置的生化反应区, 经过一系列硝化–反硝化过程, 最终处理后的污水经二次沉淀池沉淀后排入受纳水体(河流、湖泊等). 生化反应装置是污水处理过程的核心环节, 其主要由厌氧池和好氧池两大部分组成(如图1所示).

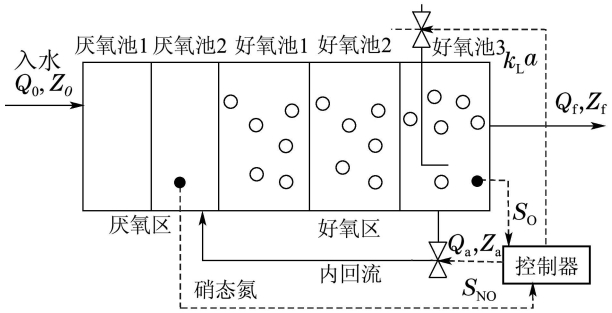

图 1 污水处理装置生化反应过程示意图

Fig. 1 Schematic diagram of biochemical process of wastewater treatment plant

有机氮的去除是生化反应过程的主要目标之一. 好氧区溶解氧溶度($S_{\rm O}$)及厌氧区硝态氮浓度($S_{\rm NO}$)对有机氮的去除效果有重要影响. $S_{\rm O}$的水平过低, 会使污泥活性降低, 抑制生物对有机物的降解, 易产生污泥膨胀; 而$S_{\rm O}$浓度过高会加速消耗污水中的有机物, 增加系统能耗[3]. $S_{\rm NO}$浓度过低, 会影响反硝化过程的速度, 而其浓度过高则会增加出水硝态氮的含量, 进而增加后续处理成本. 所以, 实现$S_{\rm O}$及$S_{\rm NO}$的平稳控制对于提高出水质量至关重要.

3 启发式动态规划控制器设计(Design of the heuristic dynamic programming controller)

污水处理过程可表示为

$$\boldsymbol{x}(k+1)=F(\boldsymbol{x}(k),\boldsymbol{u}(k)), \tag{1}$$

其中: $\boldsymbol{x}(k)$是系统状态, $\boldsymbol{u}(k)$为控制量, 其维数分别为n, m, $F(\cdot)$是关于$\boldsymbol{x}(k)$和$\boldsymbol{u}(k)$的未知非线性函数. 这里$\boldsymbol{u}(k)$是系统状态$\boldsymbol{x}(k)$的函数, 即

$$\boldsymbol{u}(k)=h(\boldsymbol{x}(k)). \tag{2}$$

系统(1)的最优控制问题可归结为选择合适的控制策略$\boldsymbol{u}$实现式(3)的性能指标最小化[5].

$$J(\boldsymbol{x}(k))=\sum_{j=k}^{\infty}\gamma^{j-k}U(\boldsymbol{x}(j)), \tag{3}$$

其中: $J(\boldsymbol{x}(k))$为评价函数, $U(\boldsymbol{x}(k))$为立即评价, 即一个周期内引发的成本值, $0<\gamma\leqslant 1$是折扣因子. 为方便起见, $J(\boldsymbol{x}(k))$简记为$J(k)$, $U(\boldsymbol{x}(k))$则简记为$U(k)$. 式(3)也可写为

$$J(k)=U(k)+\gamma J(k+1). \tag{4}$$

根据Bellman优化原理, 最优的控制策略为[7]

$$\boldsymbol{u}^*(k)=\arg\min_{u}\left[U(k)+\gamma J(k+1)\right]. \tag{5}$$

HDP采用逐步逼近的方式对式(5)进行求解. 其原理如图2所示.

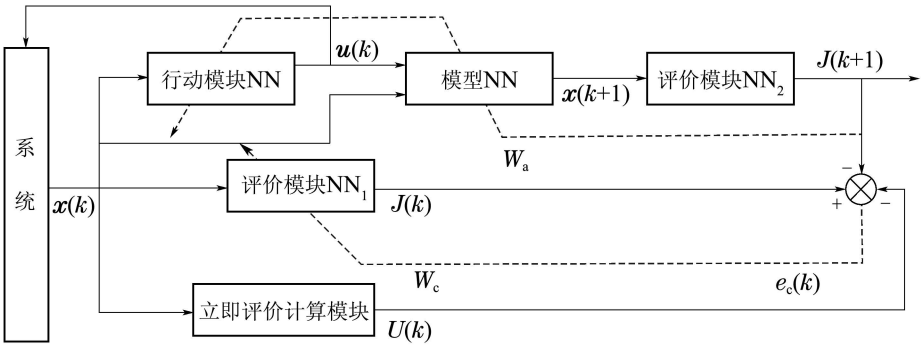

图 2 HDP控制器结构(虚线表示评价网络及行动网络的权值调整路径)

Fig. 2 Architecture of HDP controller (the dotted lines correspond to the learning path of the cctor NN and the critic NN)

HDP控制器中, 策略评价和策略提升分别由评价模块(critic)和行动模块(actor)完成. 这两个模块采用神经网络实现. 其中: actor NN为控制网络; critic $\rm NN_1$与critic $\rm NN_2$为评价网络, 其结构及参数均相同, 但是critic $\rm NN_1$的输入和输出分别为$\boldsymbol{x}(k)$及$J(k)$, 而critic $\rm NN_2$的输入和输出分别为$\boldsymbol{x}(k+1)$及$J(k+1)$; 模型NN为辨识网络, 其功能是对系统下一步的状态进行预测; utility为立即评价计算模块. 这里所有的神经网络均采用回声状态网络(echo state network, ESN).

3.1 回声状态网络(Echo state network)

ESN是Jaeger于2001年提出的一种新型递归神经网络[11]. 其核心是一个动态神经元池(dynamic neurons reservoir, DNR), 神经元池内包含大量神经元, 这些神经元之间采用随机、稀疏的方式连接. 其结构如图3所示.

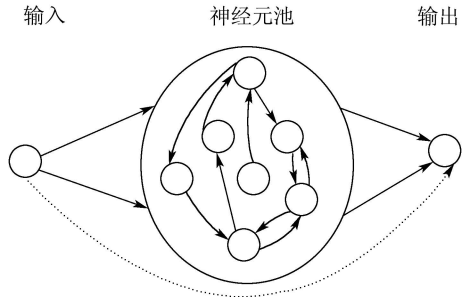

图 3 ESN结构

Fig. 3 Schematic diagram of ESN structure

ESN的隐层神经元输出可表示为

$$\boldsymbol{s}(k+1)=f(W_{\rm IN}\boldsymbol{u}(k+1)+W_{\rm R}\boldsymbol{s}(k)),\quad (6)$$

其中: $\boldsymbol{s}(k)=[\boldsymbol{s}_1(k)\ \cdots\ \boldsymbol{s}_N(k)]^{\rm T}$为内部状态, 即ESN隐层神经元输出, $\boldsymbol{u}(k)=[\boldsymbol{u}_1(k)\ \cdots\ \boldsymbol{u}_K(k)]^{\rm T}$为ESN输入. $W_{\rm IN}$, $W_{\rm R}$分别为ESN输入到隐层及隐层神经元之间的连接权值矩阵, K, N分别为输入和隐层的神经元个数.

ESN的输出可表示为

$$\boldsymbol{y}(k+1)=W^{\rm T}\boldsymbol{s}(k+1),\quad (7)$$

这里: $\boldsymbol{y}(k)=[\boldsymbol{y}_1(k)\ \cdots\ \boldsymbol{y}_L(k)]^{\rm T}$为ESN的输出, L为输出神经元个数, W为隐层到输出的连接权值矩阵. ESN中, $W_{\rm IN}$, $W_{\rm R}$随机确定, 并在网络学习过程中保持不变[11], 也就是说, ESN只需对W进行学习[12], 所以, ESN各输出对应的权值调整是独立的.

3.2 HDP控制器的学习(Learning of HDP controller)

HDP控制器的设计过程实质上是评价和控制网络通过权值参数的调整逐渐逼近最优策略的过程.

3.2.1 评价网络的学习(Learning of critic network)

设$W_{\rm c}(k)$为评价网络的可调权值, 则$W_{\rm c}$的调整可通过最小化$E_{\rm c}$进行,

$$E_{\rm c}(k)=\frac{1}{2}e_{\rm c}^2(k),\quad (8)$$

其中

$$e_{\rm c}(k)=J(k)-U(k)-\gamma J(k+1).\quad (9)$$

当$E_{\rm c}(k)=0$时, 根据式(9)可知, $J(k)=U(k)-\gamma J(k+1)$成立. 这意味着当$E_{\rm c}(k)=0$时, 式(9)与式(3)等价, 即评价网络的输出即为当前策略下的成本函数值. 这说明当评价网络的训练达到稳态时, 评价网络可实现状态到评价值的映射.

根据梯度下降算法, 评价网络的权值增量为

$$\Delta W_{\rm c}(k)=-\eta_{\rm c}(k)(\frac{\partial E_{\rm c}(k)}{\partial W_{\rm c}(k)})^{\rm T}=$$

$$-\eta_{\rm c}(k)e_{\rm c}(k)(\frac{\partial e_{\rm c}(k)}{\partial W_{\rm c}(k)})^{\rm T}=$$

$$-\eta_{\rm c}(k)e_{\rm c}(k)(\frac{\partial J(k)}{\partial W_{\rm c}(k)})^{\rm T},\quad (10)$$

这里$\eta_{\rm c}(k)$为评价网络的学习率. 因为$J(k)$是评价网络的输出, 根据式(7), 可知

$$\frac{\partial J(k)}{\partial W_{\rm c}(k)}=\boldsymbol{s}_{\rm c}^{\rm T}(k),\quad (11)$$

$\boldsymbol{s}_{\rm c}(k)$为评价网络隐层神经元的输出. 将式(11)代入式(10)可得

$$\Delta W_{\rm c}(k)=-\eta_{\rm c}(k)e_{\rm c}(k)\boldsymbol{s}_{\rm c}(k).\quad (12)$$

3.2.2 控制网络的学习(Learning of actor netork)

控制网络权值调整的目标是最小化成本函数$J(k)$, 所以其性能指标可设定为[6]

$$\frac{\partial J(k)}{\partial W_{\rm a}(k)}=\frac{\partial U(k)}{\partial W_{\rm a}(k)}+\gamma\frac{\partial J(k+1)}{\partial W_{\rm a}(k)}=0,\quad (13)$$

其中$W_{\rm a}(k)$为控制网络的可调权值. 由于各控制量对应的权值调整是独立的, 所以, 仅以控制量$\boldsymbol{u}_i$对应的权值向量$W_{{\rm a},i}$为例说明控制网络的权值调整过程. 根据梯度下降算法及链式求导法则, 控制网络的权值增量为

$$\Delta W_{{\rm a},i}(k)=-\eta_{{\rm a},i}(k)\Xi_i(k)(\frac{\partial \boldsymbol{u}_i(k)}{\partial W_{{\rm a},i}(k)})^{\rm T},\quad (14)$$

这里

$$\Xi_i(k)=\frac{\partial U(k)}{\partial \boldsymbol{u}_i(k)}+\gamma\frac{\partial J(k+1)}{\partial \boldsymbol{u}_i(k)}.\quad (15)$$

因为$\boldsymbol{u}_i(k)$是评价网络的输出, 根据式(7), 有

$$\frac{\partial \boldsymbol{u}_i(k)}{\partial W_{{\rm a},i}(k)}=\boldsymbol{s}_{\rm a}^{\rm T}(k),\quad (16)$$

$\boldsymbol{s}_{\rm a}(k)$为控制网络隐层神经元的输出. 将式(16)代入式(14)可得

$$\Delta W_{{\rm a},i}(k)=-\eta_{{\rm a},i}(k)\Xi_i(k)\boldsymbol{s}_{\rm a}(k).\quad (17)$$

3.3 控制器的收敛性分析(Convergence analysis of HDP controller)

定义离散的Lyapunov函数为

$$L(k)=\frac{1}{2}e^2(k),\quad (18)$$

$e(k)$为与权值调整过程相应的误差性能指标, 则

$$\Delta L(k)=\frac{1}{2}e^2(k+1)-\frac{1}{2}e^2(k)=$$

$$\Delta e(k)[e(k)+\frac{1}{2}\Delta e(k)],\quad (19)$$

这里$\Delta e(k)=e(k+1)-e(k)$. 设W为神经网络权值, 根据全微分定理, 有

$$\Delta e(k)=\frac{\partial e(k)}{\partial W(k)}\Delta W(k).\quad (20)$$

假设评价及控制网络的权值有界. 即存在正数$w_{\rm cM}$及$w_{\rm aM}$, 使$\|W_{\rm c}(k)\|<w_{\rm cM}$, $\|W_{\rm a}(k)\|<w_{\rm aM}$.

定理 1　若$\eta_{\rm c}(k)$满足

$$\eta_{\rm c}(k) < \frac{2}{\|\boldsymbol{s}_{\rm c}(k)\|^2}, \tag{21}$$

则评价网络的权值学习过程是收敛的.

证　定义$e(k)=e_{\rm c}(k)$, $W(k)=W_{\rm c}(k)$. 根据式(9)(11), 有

$$\frac{\partial e(k)}{\partial W(k)} = \frac{\partial J(k)}{\partial W(k)} = \boldsymbol{s}_{\rm c}^{\rm T}(k). \tag{22}$$

将式(12)及式(22)代入式(20),可得

$$\Delta e(k) = -\eta_{\rm c}(k)e(k)\|\boldsymbol{s}_{\rm c}(k)\|^2, \tag{23}$$

将式(23)代入式(19), 可得

$$\Delta L(k) = -\eta_{\rm c}(k)e^2(k)\|\boldsymbol{s}_{\rm c}(k)\|^2 \cdot [1-\frac{1}{2}\eta_{\rm c}(k)\|\boldsymbol{s}_{\rm c}(k)\|^2], \tag{24}$$

所以只要

$$1-\frac{1}{2}\eta_{\rm c}(k)\|\boldsymbol{s}_{\rm c}(k)\|^2 > 0, \tag{25}$$

则$\Delta L(k)\leqslant 0$, 并且仅当$e(k)=0$时, $\Delta L(k)=0$. 解不等式(25)即可得到式(21). 根据离散系统的Lyapunov理论, 当式(21)成立时, 评价网络值学习过程是收敛的.　　证毕.

定理 2　若$\eta_{{\rm a},i}(k)$满足

$$\eta_{{\rm a},i}(k) < \frac{2(U(k)+\gamma J(k+1))}{\Xi_i^2(k)\|\boldsymbol{s}_{\rm a}(k)\|^2}, \tag{26}$$

则控制网络的权值调整过程是收敛的.

证　令

$$e(k)=U(k)+\gamma J(k+1),\ W(k)=W_{{\rm a},i}(k),$$

则

$$\frac{\partial e(k)}{\partial W(k)} = \frac{\partial U(k)}{\partial W_{{\rm a},i}(k)} + \gamma\frac{\partial J(k+1)}{\partial W_{{\rm a},i}(k)}. \tag{27}$$

根据链式偏导法则, 并将式(16)代入式(27), 可得

$$\frac{\partial e(k)}{\partial W(k)} = \Xi_i(k)\boldsymbol{s}_{\rm a}^{\rm T}(k), \tag{28}$$

将式(17)(28)代入式(20)可得

$$\Delta e(k) = -\eta_{{\rm a},i}(k)\Xi_i^2(k)\|\boldsymbol{s}_{\rm a}(k)\|^2, \tag{29}$$

将式(29)代入式(19)可得

$$\Delta L(k) = -\eta_{{\rm a},i}(k)\Xi_i^2(k)\|\boldsymbol{s}_{\rm a}(k)\|^2 [e(k)-\frac{1}{2}\eta_{{\rm a},i}(k)\Xi_i^2(k)\|\boldsymbol{s}_{\rm a}(k)\|^2]. \tag{30}$$

当满足

$$e(k) > \frac{1}{2}\eta_{{\rm a},i}(k)\Xi_i^2(k)\|\boldsymbol{s}_{\rm a}(k)\|^2, \tag{31}$$

$\Delta L(k)\leqslant 0$. 解不等式(31), 即可得到式(26).　证毕.

值得说明的是, 与评价网络不同, 在控制网络的学习过程中, $e(k)=U(k)+\gamma J(k+1)$并不一定趋向于0(这取决于立即评价的定义方式), 而是趋向于一个极小值. 从式(30)可以看出, 在满足式(26)条件下, 仅当 $\Xi_i(k)=0$ 时, $\Delta L(k)=0$. 而$\Xi_i(k)=0$正是控制网络学习的目标(见式(13)).

4　实验研究(Experimental studies)

由于不同污水处理过程的入水情况不同, 这使得控制策略比较十分困难. 为此, 国际水协会提出了一个用于污水处理过程控制策略测试及比较的标准平台(benchmark simulation no. 1, BSM1). BSM1对污水处理过程装置的标准模型、入水条件、控制策略的测试方法及控制策略的评价标准都进行了明确的定义, 该平台能够实现不同控制策略在相同条件下的客观比较. 所以, 本文所有实验均基于BSM1进行.

4.1　立即评价设计(Design of utility)

HDP控制器设计中, 首先要确定立即评价指标$U(k)$的形式. 对于污水处理过程的跟踪控制而言, 溶解氧及硝态氮浓度的控制精度对于出水质量有较大影响. 其控制目标是使溶解氧和硝态氮浓度尽可能接近相应的设定值, 所以, 将$U(k)$的形式定义为

$$U(k) = \frac{1}{2}\boldsymbol{E}^{\rm T}(k)\boldsymbol{E}(k), \tag{32}$$

这里:

$$\boldsymbol{E}(k) = [E_1(k)\ \ E_2(k)],$$
$$E_1(k) = R_{S_{\rm O}}(k) - S_{\rm O}(k),$$
$$E_2(k) = R_{S_{\rm NO}}(k) - S_{\rm NO}(k),$$
$$R_{S_{\rm O}} = 2\,{\rm mg/L},\ R_{S_{\rm NO}} = 1\,{\rm mg/L}$$

分别为溶解氧及硝态氮浓度的设定值[13].

4.2　参数初始化(Parameter initialization)

评价网络的输入为$\boldsymbol{x}(k)=[S_{\rm O}(k)\ \ S_{\rm NO}(k)]^{\rm T}$, 输出为$J(k)$; model ESN的输入为$[K_{\rm L}a(k)\ \ Q_{\rm a}(k)\ \ S_{\rm O}(k)\ \ S_{\rm NO}(k)]^{\rm T}$, 控制网络的输入为

$$[K_{\rm L}a(k)\ \ Q_{\rm a}(k)\ \ S_{\rm O}(k)\ \ S_{\rm NO}(k)\ \ X_{\rm BH}(k)\ \ X_{\rm BA}(k)\ \ Q_{\rm in}(k)]^{\rm T},$$

输出为 $[\Delta K_{\rm L}a(k)\ \ \Delta Q_{\rm a}(k)]^{\rm T}$. $K_{\rm L}a(k)$ 为好氧区的空气转换系数, $Q_{\rm a}(k)$ 为内回流流量, 而 $X_{\rm BH}(k)$, $X_{\rm BA}(k)$和$Q_{\rm in}(k)$分别为入水异养菌浓度、入水自养菌浓度以及入水流量. 系统的采样周期$T=1.25\times10^{-2}\,{\rm h}=45\,{\rm s}$; 评价网络, 控制网络及辨识网络的隐层神经元个数分别为40, 40和50; 其谱半径[11]约为0.68, 0.68 和0.76; 折扣因子$\gamma=0.95$. 实验主要将HDP与PID控制器进行对比, PID控制器的参数

为: 对于溶解氧控制回路, $K_{\rm P}=200$, $T_{\rm I}=15$, $T_{\rm D}=2$; 对于硝态氮控制回路,

$$K_{\rm P}=20000,\ T_{\rm I}=5000,\ T_{\rm D}=400.$$

4.3 实验结果及分析(Result and analysis)

4.3.1 解耦能力测试(Decoupling ability test)

首先对HDP的解耦能力进行测试, 将入水流量固定在$Q_{\rm in}=21471\ {\rm m^3/d}$, 溶解氧及硝态氮浓度的设定值按照如下的规则调整:

$$\begin{cases} R_{S_{\rm O}}=2,\ R_{S_{\rm NO}}=1, & t<0.5,\\ R_{S_{\rm O}}=2,\ R_{S_{\rm NO}}=0.8, & 0.5\leqslant {\rm t}<0.75,\\ R_{S_{\rm O}}=1.9,\ R_{S_{\rm NO}}=0.8, & t\geqslant 0.75, \end{cases} \tag{33}$$

式中t表示运行时间(d). 在被控量设定值按照上述规律改变时, HDP及PID的控制效果如图4所示.

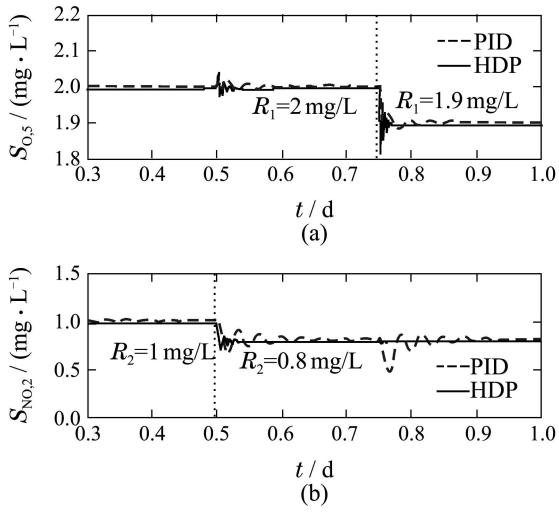

图 4 HDP的解耦性能

Fig. 4 Decoupling ability of HDP controller

由图4可看出, 当$R_{S_{\rm NO}}$从1 mg/L变为0.8 mg/L时(图4(b)), $S_{\rm O}$产生了较小的波动(图4(a)), 这说明硝态氮浓度变化对好氧区溶解氧浓度影响较小. 在PID控制作用下, 当$R_{S_{\rm O}}$从2 mg/L降低到1.9 mg/L时(图4(a)), $S_{\rm NO}$波动较大(图4(b)). 这说明好氧区溶解氧浓度变化对厌氧区硝态氮浓度有较大影响, 二者耦合作用明显. 在HDP控制器作用下, $R_{S_{\rm O}}$改变时, $S_{\rm NO}$只出现了轻微的波动, 这表明HDP控制器解耦能力较强.

4.3.2 控制性能测试(Control performance test)

接下来对HDP控制器的整体控制性能进行测试, BSM1提供了14天的入水数据, 这些数据均来自实际的污水处理厂. 入水流量及部分污染物浓度变化如图5所示.

图6为HDP和PID控制器的控制效果. 可以看出, HDP控制器作用下, $S_{\rm O}$及$S_{\rm NO}$的波动范围明显减小, 控制精度明显提高. 当较大的干扰出现时(第8–12天, 入水流量及入水污染物浓度发生了较大的波动), HDP的溶解氧浓度控制性能较正常时略有下降(图5(a)第8–12天), 但经过一段时间的学习, HDP的控制性能得到恢复, 这说明HDP自身具有较强的自学习能力.

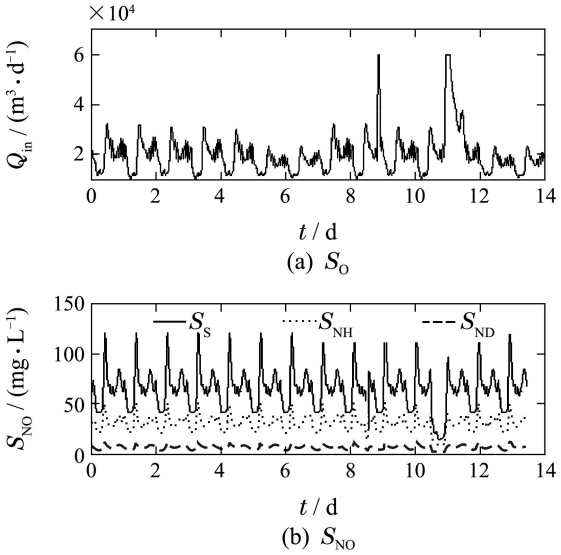

图 5 入水流量及部分入水污染物浓度

Fig. 5 Influent flow rate and partial influent pollution concentration

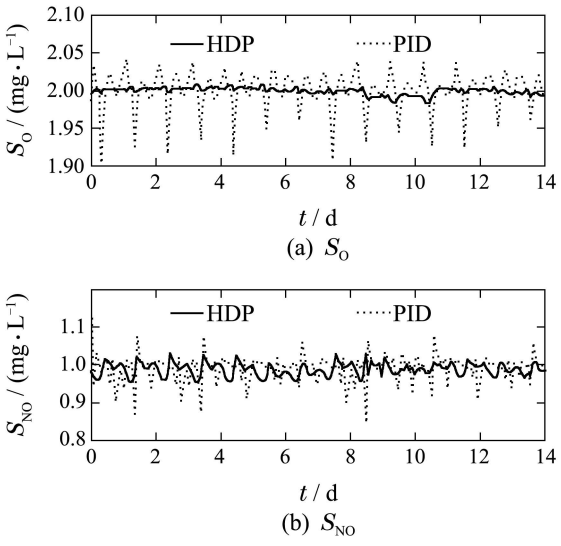

图 6 HDP与PID的控制性能比较

Fig. 6 Comparison of HDP controller and PID controller

表1列出了HDP及PID控制器的与控制精度相关的评价指标. 对于溶解氧浓度控制, HDP与PID比较, 绝对平均误差降低了73.4%, 最大偏差降低了79.1%, 误差方差降低了1个数量级. 对于硝态氮浓度控制, HDP与PID相比较, 绝对平均误差降低了56.3%, 最大偏差降低了74.4%, 误差方差降低了1个数量级. 可以看出, HDP在控制精度及控制的平稳性上明显优于PID控制器.

表 1 2种控制器的控制精度评价指标

Table 1 Evaluation indices of control precision under two controllers

控制变量	PID		HDP	
	$S_{\rm O}$	$S_{\rm NO}$	$S_{\rm O}$	$S_{\rm NO}$
绝对平均误差/$(\mathrm{mg}\cdot\mathrm{L}^{-1})$	0.0139	0.0268	0.0037	0.0117
绝对最大误差/$(\mathrm{mg}\cdot\mathrm{L}^{-1})$	0.0980	0.1756	0.0204	0.0450
误差方差/$(\mathrm{mg}\cdot\mathrm{L}^{-1})^2$	2.54×10^{-4}	0.0016	4.16×10^{-5}	2.20×10^{-4}

在HDP的在线学习过程中, 学习率的选择为

$$\begin{cases}\eta_{\rm c}(k)=\dfrac{1}{\|\boldsymbol{s}_{\rm c}(k)\|^2},\\ \eta_{{\rm a},i}(k)=\dfrac{U(k)+\gamma J(k+1)}{\Xi_i^2(k)\|\boldsymbol{s}_{\rm a}(k)\|^2+\epsilon},\end{cases}\tag{34}$$

这里ϵ为一小的正数.

5 结论(Conclusions)

溶解氧和硝态氮浓度是污水处理过程的两个关键指标, 由于实际污水处理过程的生化反应模型难以建立, 并且两个变量之间存在严重的耦合, 所以其控制难度较大. 启发式动态规划采用策略评价和策略提升的方法逐步逼近最优的控制策略, 在策略搜索的过程中HDP能够综合考虑各指标之间的耦合关系, 所以HDP控制器具有较强的解耦能力. 此外, HDP不需要污水处理的生化反应过程的机理模型, 其通过对输入输出的测量数据的学习即可实现策略评价和策略提升. 即HDP能够以数据驱动的模式运行. 对于污水处理过程这类机理模型难以建立的复杂系统, HDP实现简单, 有较好的应用前景.

参考文献(References):

[1] 史雄伟, 乔俊飞, 苑明哲. 基于改进粒子群优化算法的污水处理过程优化控制 [J]. 信息与控制, 2011, 40(5): 698 – 703.
(SHI Xiongwei, QIAO Junfei, YUAN Mingzhe. Optimal control for wastewater treatment process based on improved particle swarm optimization algorithm [J]. *Information and Control*, 2011, 40(5): 698 – 703.)

[2] SHEN W H, CHEN X Q, PONS M N. Model predictive control for wastewater treatment process with feedforward compensation [J]. *Chemical Engineering Journal*, 2009, 155(1/2): 161 – 174.

[3] 丛秋梅, 柴天佑, 余文. 污水处理过程的递阶神经网络建模 [J]. 控制理论与应用, 2009, 26(1): 8 – 14.
(CONG Qiumei, CHAI Tianyou, YU Wen. Modeling wastewater treatment plant via hierarchical neural networks [J]. *Control Theory & Applications*, 2009, 26(1): 8 – 14.)

[4] DELLANA S A, WEST D. Predictive modeling for wastewater applications: Linear and nonlinear approaches [J]. *Environmental Modelling & Software*, 2009, 24(1): 96 – 106.

[5] WANG F Y, ZHANG H G, LIU D R. Adaptive dynamic programming: an introduction [J]. *IEEE Computational Intelligence Magazine*, 2009, 4(2): 39 – 47.

[6] PROKHOROV D V, WUNSCH D C. Adaptive critic designs [J]. *IEEE Transactions on Neural Networks*, 1997, 8(5): 997 – 1007.

[7] LEWIS F L, VRABIE D. Reinforcement learning and adaptive dynamic programming for feedback control [J]. *IEEE Circuits and Systems Magzine*, 2009, 9(3): 32 – 50.

[8] VENAYAGAMOORTHY G K, HARLEY R G, WUNSCH D C. Comparison of a heuristic dynamic programming and a dual heuristic programming based adaptive critics neurocontroller for a turbogenerator [C] //*Proceedings of the International Joint Conference on Neural Networks*. Como, Italy: IEEE, 2000: 233 – 238.

[9] 曹宁, 张化光, 罗艳红, 等. 一类非线性奇异摄动系统的近似最优控制 [J]. 控制理论与应用, 2011, 28(5): 688 – 692.
(CAO Ning, ZHANG Huaguang, LUO Yanhong, et al. Suboptimal control of a class of nonlinear singularly perturbed systems [J]. *Control Theory & Applications*, 2011, 28(5): 688 – 692.)

[10] ZHAO D B, BAI X R, WANG F Y, et al. DHP method for ramp metering of freeway traffic [J]. *IEEE Transactions on Intelligent Transportation System*, 2011, 12(4): 990 – 999.

[11] JAEGER H. Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication [J]. *Science*, 2004, 304(5667): 78 – 80.

[12] 韩敏, 穆大芸. 基于Adaboost算法的回声状态网络预报器 [J]. 控制理论与应用, 2011, 28(4): 601 – 604.
(HAN Min, MU Dayun. Improvement of echo state network accuracy with adaboost [J]. *Control Theory & Applications*, 2011, 28(4): 601 – 604.)

[13] ALEX J, MAGDEBURG V. *Benchmark Simulation Model No. 1 (BSM1)* [S]. IWA Taskgroup on Benchmarking of Control Stategies for WWTPs, 2008.

作者简介:

薄迎春　(1977—), 男, 博士研究生, 目前研究方向为神经计算及智能优化控制, E-mail: boyingchun@sina.com.cn;

乔俊飞　(1968—), 男, 教授, 博士生导师, 目前研究方向为复杂系统建模、智能优化控制、神经网络等, E-mail: isbox@sina.com.cn.