基于 ESN 的多指标 DHP 控制策略在污水处理过程中的应用

乔俊飞 [1] 薄迎春 [1, 2] 韩 广 [1]

摘 要 针对污水处理过程 (Wastewater treatment process, WWTP) 溶解氧 (Dissolved oxygen, DO) 及硝态氮浓度控制问题, 提出了一种多评价指标的 DHP (Dual heuristic dynamic programming) 控制策略. 该策略能够降低评价指标的复杂性, 提高评价网络的逼近精度. 采用回声状态网络 (Echo state networks, ESNs) 实现评价函数及控制策略的逼近, 研究了控制器的在线学习算法. 实验表明, 该策略在控制性能上优于单评价指标的 DHP 策略及常规 PID 控制策略.

关键词 自适应动态规划, 多评价指标, 污水处理, 回声状态网络

引用格式 乔俊飞, 薄迎春, 韩广. 基于 ESN 的多指标 DHP 控制策略在污水处理过程中的应用. 自动化学报, 2013, **39**(7): 1146−1151

DOI 10.3724/SP.J.1004.2013.01146

Application of ESN-based Multi Indices Dual Heuristic Dynamic Programming on Wastewater Treatment Process

QIAO Jun-Fei[1] BO Ying-Chun[1, 2] HAN Guang[1]

Abstract In order to solve the problem of controlling dissolved oxygen (DO) concentration and nitrate concentration of wastewater treatment process (WWTP), a multi critic indices dual heuristic dynamic programming (MDHP) policy is proposed. The approximating precision can be improved through lowering the complexity between the relationship of the critic network's outputs and inputs in this scheme. Echo state networks (ESNs) are adopted to approximate the critic indices and the optimal control policy. Online learning method of the controller is investigated. Experimental results indicate that the MDHP scheme has some advantages over single critic index DHP (SDHP) and PID in control performance.

Key words Adaptive dynamical programming (ADP), multi evaluation indices, wastewater treatment, echo state network (ESN)

Citation Qiao Jun-Fei, Bo Ying-Chun, Han Guang. Application of ESN-based multi indices dual heuristic dynamic programming on wastewater treatment process. *Acta Automatica Sinica*, 2013, **39**(7): 1146−1151

好氧区溶解氧浓度及厌氧区硝态氮浓度是活性污泥法污水处理过程的重要指标[1−2]. 好氧区溶解氧浓度过低, 会抑制生物对有机物的降解, 易产生污泥膨胀; 而其浓度过高会加速消耗污水中的有机物, 降低活性污泥的絮凝性能和吸附能力, 增加能耗[1]. 厌氧区硝态氮浓度过低, 会减缓反硝化过程的速度, 而其浓度过高则会影响磷的去除. 此外, 厌氧区硝态氮浓度受控于好氧区到厌氧区的回流流量, 所以好氧区溶解氧浓度对厌氧区硝态氮浓度影响较大, 二者存在较强的耦合关系. 针对溶解氧及硝态氮浓度控制问题, 模型预测控制技术近年来得到了广泛的研究[2]. 然而, 模型预测控制主要以污水处理过程的机理模型为基础, 由于污水处理过程参数的时变特性及生化反应过程的不确定性, 适应于控制需求的机理模型构造仍是一个未解决的问题[3].

针对一类模型难以确定的被控过程, 自适应动态规划 (Adaptive dynamic programming, ADP) 在近几年得到了一定的重视[4−11]. 其最大的优点是控制器的设计对系统的动力学模型依赖较小. 典型的 ADP 采用 "评价 – 行动" 迭代的方式逐步逼近最优的控制策略. DHP (Dual heuristic dynamic programming)[11] 是实现 ADP 的主要方法之一, 已经在复杂非线性系统控制领域得到了广泛的研究和应用[7−10].

目前的 DHP 设计只考虑单一的评价指标, 显然, 如果能够将评价指标分解成多个分量, 则可以降低评价指标的复杂性, 同时, 可以更清晰地了解每个控制量对每个指标分量的影响, 这对于提高评价精度及控制器的策略搜索能力非常有利[11]. 基于这个思想, 本文提出了一种基于回声状态网络 (Echo state network, ESN)[12] 的多指标 DHP (E-MDHP) 控制策略, 并将其应用在污水处理过程的溶解氧和硝态氮浓度控制中.

1 问题描述

一般的优化问题均设定一性能指标函数, 形式如下[5]:

$$Q_h(\boldsymbol{x}(k),\boldsymbol{u}(k)) = \sum_{i=k}^{\infty} \gamma^{i-k} r(\boldsymbol{x}(i),\boldsymbol{u}(i)) \tag{1}$$

其中, $Q_h(\boldsymbol{x}(k),\boldsymbol{u}(k))$ 为评价函数, $r(\boldsymbol{x}(k),\boldsymbol{u}(k))$ 为立即评价, $0 < \gamma \leq 1$ 为评价因子. $\boldsymbol{x}(k)$ 表示系统的状态, $\boldsymbol{u}(k)$ 为控制策略. 为方便起见, $Q_h(\boldsymbol{x}(k),\boldsymbol{u}(k))$ 简记为 $Q(k)$, $r(\boldsymbol{x}(k),\boldsymbol{u}(k))$ 简记为 $r(k)$. 式 (1) 亦可写为如下形式:

$$Q(k) = r(k) + \gamma Q(k+1) \tag{2}$$

优化的目的是使性能指标 Q 最小化, 根据 Bellman 优化原理, 最优的控制策略为

$$\boldsymbol{u}(k) = \arg\min_{\boldsymbol{u}}(Q(k)) \tag{3}$$

ADP 采用迭代的方式逐步逼近式 (3) 的最优解. DHP 中, 评价网络输出为 Q 对 $\boldsymbol{x}$ 的导数. 在常规 DHP 中, Q 是一个标量, 即 $\boldsymbol{\lambda}(k)$ 的维数为 $\boldsymbol{x}$ 的维数 (记为 N_x). 一般情况下, 多变量控制中的性能指标可以分解为多个分量, 即: $\boldsymbol{Q} = [\boldsymbol{Q}_1, \cdots, \boldsymbol{Q}_{N_q}]^{\mathrm{T}}$ (N_q 为指标分量个数). 此时

$$\boldsymbol{\lambda}(k) = \begin{bmatrix} \dfrac{\partial \boldsymbol{Q}_1(k)}{\partial \boldsymbol{x}_1(k)} & \cdots & \dfrac{\partial \boldsymbol{Q}_1(k)}{\partial \boldsymbol{x}_{N_x}(k)} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial \boldsymbol{Q}_{N_q}(k)}{\partial \boldsymbol{x}_1(k)} & \cdots & \dfrac{\partial \boldsymbol{Q}_{N_q}(k)}{\partial \boldsymbol{x}_{N_x}(k)} \end{bmatrix} \tag{4}$$

其维数为 $N_q \times N_x$. 指标分解有如下优点: 1) 分解后的各指标分量与输入间的关系更为简单, 文献 [11] 指出, 如果一个神经网络的输出可以分解为多个分量, 则通过单独训练各分量可以得到更好的逼近精度. 所以, 指标分解为提高 $\boldsymbol{\lambda}$ 的逼

收稿日期 2011-12-30 录用日期 2012-11-29
Manuscript received December 30, 2011; accepted November 29, 2012
国家自然科学基金 (61034008), 教育部博士点基金 (200800050004), 北京市自然科学基金 (4092010) 资助
Supported by National Natural Science Foundation of China (61034008), Doctoral Fund of Ministry of Education of China (200800050004) and Beijing Municipal Natural Science Foundation (4092010)
本文责任编委 储健
Recommended by Associate Editor CHU Jian
1. 北京工业大学电子信息与控制工程学院 北京 100124 2. 中国石油大学信息与控制工程学院 青岛 266580
1. College of Electronic and Control Engineering, Beijing University of Technology, Beijing 100124 2. College of Information and Control Engineering, China University of Petroleum, Qingdao 266580

近精度提供了可能性. 2) 分解后的各指标分量与各控制量之间的关系更为清晰, 有利于将已知的先验知识加入到控制器的设计中.

但是, 这种做法的前提是每个输出分量对应的参数可以独立学习, 传统的神经网络 (如 BP 网络及传统递归神经网络) 各输出共用隐层的权值, 所以各输出对应的权值无法通过一个神经网络实现独立训练. ESN 是一种池计算 (Reservoir computing, RC)[12] 神经网络, 其优点是只需要训练隐层与输出之间的连接权值, 即 ESN 各输出相应的权值训练是相互独立的, 所以 ESN 从本质上容易解决指标分解后引起的多输出及权值独立训练问题, 因此, 选择 ESN 作为 E-MDHP 中评价及控制网络的逼近器.

2　E-MDHP 控制器设计

2.1　ESN 简介

在不考虑输出到隐层反馈的情况下, ESN 的内部状态 (隐层神经元输出) 可表示为[12]

$$\boldsymbol{s}(k+1)=f(W_{IN}\boldsymbol{u}(k+1)+W\boldsymbol{s}(k)) \tag{5}$$

其中, $\boldsymbol{s}(k)=[\boldsymbol{s}_1(k),\cdots,\boldsymbol{s}_N(k)]^{\mathrm{T}}$ 为内部状态, $\boldsymbol{u}(k)=[\boldsymbol{u}_1(k),\cdots,\boldsymbol{u}_K(k)]^{\mathrm{T}}$ 为网络输入. W_{IN}, W 分别为输入到隐层及隐层神经元之间的连接权值矩阵, 这两个矩阵随机生成, 并在网络学习过程中保持不变[12]. K, N 分别为输入和隐层的神经元个数. ESN 的输出为

$$\boldsymbol{y}(k+1)=W_o^{\mathrm{T}}\boldsymbol{s}(k+1) \tag{6}$$

其中, $\boldsymbol{y}(k)=[\boldsymbol{y}_1(k),\cdots,\boldsymbol{y}_L(k)]^{\mathrm{T}}$ 为 ESN 输出, L 为输出维数, W_o 为隐层到输出的连接权值矩阵, ESN 只需对 W_o 进行学习, 所以, 对于每个输出而言, 其对应的权值调整是独立的.

2.2　E-MDHP 控制器结构

E-MDHP 控制器结构如图 1 所示.

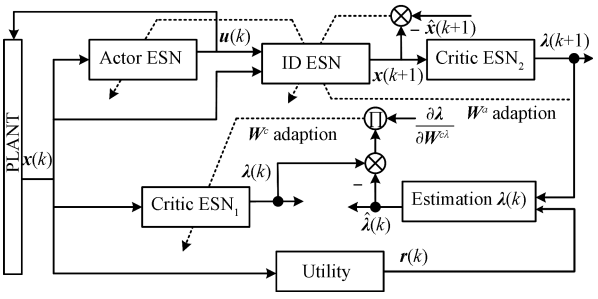

图 1　E-MDHP 结构 (虚线表示各网络的权值调整路径)

Fig. 1　E-MDHP architecture (The dotted lines are correspond to the learning path of each ESN.)

E-MDHP 共包含 4 个 ESN. 其中, Actor ESN 为控制网络; Critic ESN_1 与 Critic ESN_2 为评价网络, 这两个网络结构及参数均相同, 但 Critic ESN_1 的输入输出分别为 $\boldsymbol{x}(k)$ 及 $\boldsymbol{\lambda}(k)$, 而 Critic ESN_2 的输入输出分别为 $\boldsymbol{x}(k+1)$ 及 $\boldsymbol{\lambda}(k+1)$; ID ESN 为辨识网络, 输入为 $[\boldsymbol{x}(k),\boldsymbol{u}(k)]^{\mathrm{T}}$, 输出为 $\boldsymbol{x}(k+1)$.

2.3　E-MDHP 控制器的在线学习

E-MDHP 控制器的在线学习过程实质上是控制器中各网络权值的在线调整过程. 神经网络训练的目标是寻求一组最优权值使某一性能指标最小化. 即:

$$\boldsymbol{w}^*=\arg\min_{\boldsymbol{w}}\{F(\boldsymbol{w})\} \tag{7}$$

其中, $F(\boldsymbol{w})$ 为与训练误差相关的性能指标, 一般 $F(\boldsymbol{w})$ 形式如下:

$$F(\boldsymbol{w})=\frac{1}{2}\boldsymbol{E}^{\mathrm{T}}(\boldsymbol{w})\boldsymbol{E}(\boldsymbol{w}) \tag{8}$$

按照 Levenberg-Marquardt (LM) 算法, 权值的更新过程为

$$\Delta\boldsymbol{w}=-\left(J^{\mathrm{T}}(\boldsymbol{w})J(\boldsymbol{w})+\mu\boldsymbol{I}\right)^{-1}J^{\mathrm{T}}(\boldsymbol{w})\boldsymbol{E}(\boldsymbol{w}) \tag{9}$$

其中, μ 为一正数, $J(\boldsymbol{w})$ 为 $\boldsymbol{E}(\boldsymbol{w})$ 对 $\boldsymbol{w}$ 的雅可比矩阵. 本文采用 LM 算法实现评价网络及控制网络的训练, 显然, 应用 LM 算法的核心问题在于误差函数 $\boldsymbol{E}$ 的构造及其对权值的雅可比矩阵 J 的计算.

2.3.1　评价网络在线学习

由于 ESN 中每个输出所对应的权值是独立的, 所以仅以 $\boldsymbol{\lambda}_{ij}$ 对应的权值向量 $\boldsymbol{W}^{c_{ij}}$ 为例, 说明评价网络的权值调整过程. 评价网络的目标是对当前控制策略进行评价, 进而为控制策略的调整提供依据. 其相应的性能指标可设为[5]

$$F^{c_{ij}}(k)=\frac{1}{2}\left(\boldsymbol{E}^{c_{ij}}(k)\right)^2 \tag{10}$$

$\boldsymbol{E}^{c_{ij}}(k)$ 可定义为

$$\boldsymbol{E}^{c_{ij}}(k)=\boldsymbol{\lambda}_{ij}(k)-\hat{\boldsymbol{\lambda}}_{ij}(k) \tag{11}$$

其中

$$\hat{\boldsymbol{\lambda}}_{ij}(k)=\gamma\frac{\partial\boldsymbol{Q}_i(k+1)}{\partial\boldsymbol{x}_j(k)}+\frac{\partial\boldsymbol{r}_i(k)}{\partial\boldsymbol{x}_j(k)} \tag{12}$$

$\hat{\boldsymbol{\lambda}}_{ij}(k)$ 实质上是当前策略下 $\boldsymbol{\lambda}_{ij}(k)$ 所对应的期望值的估计[13]. $\boldsymbol{\lambda}_{ij}(k)$ 为 Critic ESN_1 的输出, 根据式 (6)

$$\frac{\partial\boldsymbol{\lambda}_{ij}(k)}{\partial\boldsymbol{W}^{c_{ij}}(k)}=\left(\boldsymbol{s}^c(k)\right)^{\mathrm{T}} \tag{13}$$

这里, $\boldsymbol{s}^c(k)$ 为 Critic ESN_1 的内部状态. 所以, 相应的雅可比矩阵为

$$J^c(k)=\frac{\partial\boldsymbol{E}^{c_{ij}}(k)}{\partial\boldsymbol{W}^{c_{ij}}(k)}=\frac{\partial\boldsymbol{\lambda}_{ij}(k)}{\partial\boldsymbol{W}^{c_{ij}}(k)}=\left(\boldsymbol{s}^c(k)\right)^{\mathrm{T}} \tag{14}$$

将其代入式 (9), 可得:

$$\Delta\boldsymbol{W}^{c_{ij}}(k)=-\eta^{c_{ij}}(k)\left(\Xi^{c_{ij}}(k)\right)^{-1}\boldsymbol{s}^c(k)\boldsymbol{E}^{c_{ij}}(k) \tag{15}$$

其中, $\eta^{c_{ij}}(k)$ 为评价网络的学习率,

$$\Xi^{c_{ij}}(k)=\boldsymbol{s}^c(k)\left(\boldsymbol{s}^c(k)\right)^{\mathrm{T}}+\mu^{c_{ij}}(k)I \tag{16}$$

为计算式 (14), 必须首先计算 $\boldsymbol{E}^{c_{ij}}(k)$. 根据式 (11), 为计算 $\boldsymbol{E}^{c_{ij}}(k)$, 必须首先得到 $\hat{\boldsymbol{\lambda}}_{ij}(k)$. $\hat{\boldsymbol{\lambda}}_{ij}(k)$ 可通过式 (12) 计算. 因为 $\boldsymbol{r}(k)$ 与 $\boldsymbol{x}(k)$ 之间存在明确的解析关系, 而这个解析关系随具体问题而有所不同, 在 $\boldsymbol{r}(k)$ 与 $\boldsymbol{x}(k)$ 关系明确后, 易计算出 $\frac{\partial\boldsymbol{r}_i(k)}{\partial\boldsymbol{x}_j(k)}$. 对于式 (12) 右边第一项, 应用链式求导法则, 可得:

$$\begin{aligned}\frac{\partial\boldsymbol{Q}_i(k+1)}{\partial\boldsymbol{x}_j(k)}=&\sum_{m=1}^{Nx}\left(\frac{\partial\boldsymbol{Q}_i(k+1)}{\partial\boldsymbol{x}_m(k+1)}\frac{\partial\boldsymbol{x}_m(k+1)}{\partial\boldsymbol{x}_j(k)}\right)+\\&\sum_{m=1}^{Nx}\left(\frac{\partial\boldsymbol{Q}_i(k+1)}{\partial\boldsymbol{x}_m(k+1)}\sum_{n=1}^{Nu}\left(\frac{\partial\boldsymbol{x}_m(k+1)}{\partial\boldsymbol{u}_n(k)}\frac{\partial\boldsymbol{u}_n(k)}{\partial\boldsymbol{x}_j(k)}\right)\right)\end{aligned} \tag{17}$$

N_u 为控制量维数.

1) 首先

$$\frac{\partial \boldsymbol{Q}_i(k+1)}{\partial \boldsymbol{x}_m(k+1)} = \boldsymbol{\lambda}_{im}(k+1) \tag{18}$$

是 Critic ESN_2 的输出 (图 1).

2) $\boldsymbol{x}(k)$ 和 $\boldsymbol{x}(k+1)$ 分别是 ID ESN 的输入及输出, 所以 $\frac{\partial \boldsymbol{x}_m(k+1)}{\partial \boldsymbol{x}_j(k)}$ 可以通过 ID ESN 导出. 同理, $\boldsymbol{u}(k)$ 和 $\boldsymbol{x}(k+1)$ 分别是 ID ESN 的输入及输出, 所以 $\frac{\partial \boldsymbol{x}_m(k+1)}{\partial \boldsymbol{u}_n(k)}$ 也可通过 ID ESN 导出.

3) $\boldsymbol{x}(k)$ 和 $\boldsymbol{u}(k)$ 分别是 Actor ESN 的输入及输出, 所以 $\frac{\partial \boldsymbol{u}_m(k)}{\partial \boldsymbol{x}_j(k)}$ 可以通过 Actor ESN 导出.

综合 1) ~ 3), 可求得 $\partial \boldsymbol{Q}_i(k+1)/\partial \boldsymbol{u}_j(k)$, 进而将其代入式 (12) 即可求得 $\hat{\boldsymbol{\lambda}}_{ij}(k)$.

2.3.2 控制网络在线学习

对于控制网络, 设与控制量 $\boldsymbol{u}_i$ 对应的权值向量为 $\boldsymbol{W}^{a_i}$, 对于固定设定值的跟踪控制系统[5], $\boldsymbol{W}^{a_i}$ 的调整目标可定义为

$$\frac{\partial \boldsymbol{Q}(k+1)}{\partial \boldsymbol{u}_i(k)} = 0 \tag{19}$$

应用链式求导法则

$$\frac{\partial \boldsymbol{Q}(k+1)}{\partial \boldsymbol{u}_i(k)} = \boldsymbol{\lambda}(k+1)\frac{\partial \boldsymbol{x}(k+1)}{\partial \boldsymbol{u}_i(k)} \tag{20}$$

显然若 $\boldsymbol{\lambda}(k+1)=0$, 则式 (19) 成立. 所以, 控制网络调整的误差指标可定义为

$$F^a(k) = \sum_{i=1}^{N_q}\sum_{j=1}^{N_x}\boldsymbol{\lambda}_{ij}^2(k+1) \tag{21}$$

根据梯度下降算法, 权值的更新规则为

$$\Delta \boldsymbol{W}^{a_i}(k) = -\eta^{a_i}(k)\boldsymbol{s}^a(k)e^{a_i}(k) \tag{22}$$

$\boldsymbol{s}^a(k)$ 为控制网络的内部状态,

$$e^{a_i}(k) = \sum_{i=1}^{N_q}\sum_{j=1}^{N_x}\left[\boldsymbol{\lambda}_{ij}(k+1)\frac{\partial \boldsymbol{\lambda}_{ij}(k+1)}{\partial \boldsymbol{u}_i(k)}\right] \tag{23}$$

根据梯度下降算法与 LM 算法的关系可知, LM 算法下权值 $\boldsymbol{W}^{a_i}(k)$ 的调整规则为

$$\begin{aligned}\Delta \boldsymbol{W}^{a_i}(k) = &-\eta^{a_i}(k)\Big(\boldsymbol{s}^a(k)\left(\boldsymbol{s}^a(k)\right)^{\mathrm{T}} + \\ &\mu^{a_i}(k)I\Big)^{-1}\boldsymbol{s}^a(k)e^{a_i}(k)\end{aligned} \tag{24}$$

其中, $\eta^{a_i}(k)$ 为控制网络中与输出 $\boldsymbol{u}_i(k)$ 相关的权值的学习率.

2.4 控制器的收敛性分析

神经网络的学习过程的收敛性与相应的学习率密切相关. 所以, 接下来对保证评价指标及控制策略收敛的学习率取值范围进行研究.

定义离散的 Lyapunov 函数为

$$L(k) = \frac{1}{2}e^2(k) \tag{25}$$

其中, $e(k)$ 为误差指标, 则

$$\Delta L(k) = \Delta e(k)\left(e(k) + \frac{1}{2}\Delta e(k)\right) \tag{26}$$

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:

$$\Delta e(k) = \frac{\partial e(k)}{\partial W(k)}\Delta W(k) \tag{27}$$

定理 1. 若评价网络的学习率 $\eta^{c_{ij}}(k)$ 满足:

$$\eta^{c_{ij}}(k) < \frac{2}{\left(\boldsymbol{s}^c(k)\right)^{\mathrm{T}}\left(\Xi^{c_{ij}}(k)\right)^{-1}\boldsymbol{s}^c(k)} \tag{28}$$

则评价网络的学习过程是收敛的. $\Xi^{c_{ij}}(k)$ 的定义见式 (16).

证明. 评价网络与 $\boldsymbol{\lambda}_{ij}(k)$ 相关的权值 $\Delta W^{c_{ij}}$ 更新的目标是使 $\boldsymbol{E}^{c_{ij}}(k)$ 趋向于 0. 定义 $e(k) = \boldsymbol{E}^{c_{ij}}(k)$, $W(k) = W^{c_{ij}}(k)$. 根据式 (13), 有

$$\frac{\partial e(k)}{\partial W(k)} = \left(\boldsymbol{s}^c(k)\right)^{\mathrm{T}} \tag{29}$$

将式 (15) 及式 (29) 代入式 (27), 可得:

$$\Delta e(k) = -\eta^{c_{ij}}(k)\left(\boldsymbol{s}^c(k)\right)^{\mathrm{T}}\left(\Xi^{c_{ij}}(k)\right)^{-1}\boldsymbol{s}^c(k)e(k) \tag{30}$$

将式 (30) 代入式 (26), 可得:

$$\begin{aligned}\Delta L(k) = &-\eta^{c_{ij}}(k)e^2(k)\left(\boldsymbol{s}^c(k)\right)^{\mathrm{T}}\left(\Xi^{c_i}(k)\right)^{-1}\boldsymbol{s}^c(k)\times \\ &\left(1 - \frac{1}{2}\eta^{c_{ij}}(k)\left(\boldsymbol{s}^c(k)\right)^{\mathrm{T}}\left(\Xi^{c_i}(k)\right)^{-1}\boldsymbol{s}^c(k)\right)\end{aligned} \tag{31}$$

因为 $\left(\Xi^{c_{ij}}(k)\right)^{-1}$ 是正定的, 所以只要

$$1 - \frac{1}{2}\eta^{c_{ij}}(k)\left(\boldsymbol{s}^c(k)\right)^{\mathrm{T}}\left(\Xi^{c_i}(k)\right)^{-1}\boldsymbol{s}^c(k) > 0 \tag{32}$$

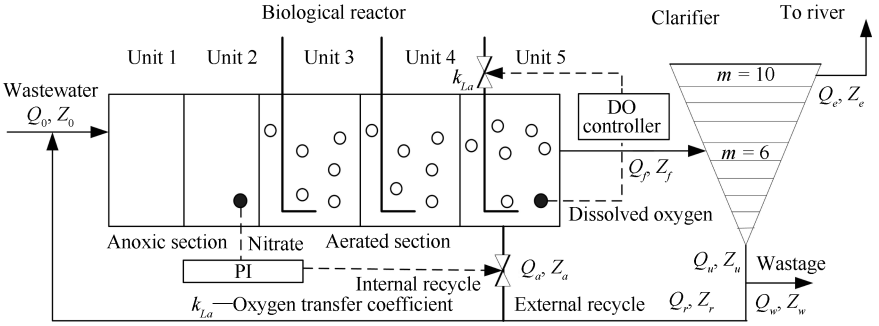

图 2 BSM1 示意图

Fig. 2 Schematic diagram of BSM1

则 $\Delta L(k) \leq 0$, 并且仅当 $e(k)=0$ 时, $\Delta L(k)=0$. 解不等式 (32) 即可得到式 (28). 根据离散系统的 Lyapunov 定理, 当式 (28) 成立时, 评价网络的学习是收敛的. □

定理 2. 若控制网络的学习率 $\eta^{a_i}(k)$ 满足

$$\eta^{a_i}(k) < \frac{\sum_{i=1}^{N_q}\sum_{j=1}^{N_x}\boldsymbol{\lambda}_{ij}^2(k+1)}{\left(e^{a_i}(k)\right)^2\left(\boldsymbol{s}^a(k)\right)^{\mathrm{T}}\left(\Xi^{a_i}(k)\right)^{-1}\boldsymbol{s}^a(k)} \tag{33}$$

则控制网络的学习过程是收敛的. 其中

$$\Xi^{a_i}(k) = \boldsymbol{s}^a(k)\left(\boldsymbol{s}^a(k)\right)^{\mathrm{T}} + \mu^{a_i}(k)I \tag{34}$$

证明. 令 $W(k)=\boldsymbol{W}^{a_i}(k)$, 并且

$$e(k) = \frac{1}{2}\sum_{i=1}^{N_q}\sum_{j=1}^{N_x}\boldsymbol{\lambda}_{ij}^2(k+1) \tag{35}$$

根据 $e(k)$ 的定义

$$\frac{\partial e(k)}{\partial W(k)} = \sum_{i=1}^{N_q}\sum_{j=1}^{N_x}\left(\boldsymbol{\lambda}_{ij}(k+1)\frac{\partial \boldsymbol{\lambda}_{ij}(k+1)}{\partial \boldsymbol{W}^{a_i}(k)}\right) \tag{36}$$

根据链式偏导法则

$$\frac{\partial \boldsymbol{\lambda}_{ij}(k+1)}{\partial \boldsymbol{W}^{a_i}(k)} = \frac{\partial \boldsymbol{\lambda}_{ij}(k+1)}{\partial \boldsymbol{u}_i(k)}\frac{\partial \boldsymbol{u}_i(k)}{\partial \boldsymbol{W}^{a_i}(k)} \tag{37}$$

$\boldsymbol{u}_i(k)$ 为控制网络的输出, 根据 ESN 的定义

$$\frac{\partial \boldsymbol{u}_i(k)}{\partial \boldsymbol{W}^{a_i}(k)} = \left(\boldsymbol{s}^a(k)\right)^{\mathrm{T}} \tag{38}$$

将式 (24), (36) 代入式 (26) 可得:

$$\Delta e(k) = -\eta^{a_i}(k)\left(e^{a_i}(k)\right)^2\left(\boldsymbol{s}^a(k)\right)^{\mathrm{T}}\left(\Xi^{a_i}(k)\right)^{-1}\boldsymbol{s}^a(k) \tag{39}$$

显然, $\Delta e(k) \leq 0$. 所以, 只要 $e(k)+\Delta e(k)/2 > 0$, 即可保证 $\Delta L(k) < 0$. 根据式 (35) 及式 (39), 可知:

$$\begin{aligned} e(k) + \frac{\Delta e(k)}{2} = {} & \frac{1}{2}\sum_{i=1}^{N_q}\sum_{j=1}^{N_x}\boldsymbol{\lambda}_{ij}^2(k+1) - \\ & \frac{1}{2}\eta^{a_i}(k)\left(e^{a_i}(k)\right)^2\left(\boldsymbol{s}^a(k)\right)^{\mathrm{T}}\left(\Xi^{a_i}(k)\right)^{-1}\boldsymbol{s}^a(k) \end{aligned} \tag{40}$$

解不等式 $e(k)+\Delta e(k)/2 > 0$, 即得到式 (33), 即当式 (33) 成立时, $\Delta L(k) < 0$. 根据离散系统的 Lyapunov 定理, 控制网络的学习过程是收敛的. □

3　实验研究

3.1　BSM1 简介

BSM1[14] 是国际水协会 (International Water Association, IWA) 提出的一个用于测试污水处理过程控制策略的标准模型 (如图 2 所示). BSM1 的控制目标将第 5 分区溶解氧的浓度 $S_{\mathrm{O},5}$ 和第二分区硝态氮的浓度 $S_{\mathrm{NO},2}$ 分别保持在 2 mg/L 和 1 mg/L. 控制量分别为第 5 分区的曝气量 K_La_5 及从第 5 分区到第 2 分区的回流流量 Q_a. 缺省的控制策略为 PID 控制策略.

BSM1 包含了三个数据文件, 分别包含了晴、雨和暴雨情况下 14 天的进水信息. 其中, 暴雨情况下的入水流量变化如图 3 所示, 从图 3 中可以看出, 由于暴雨的影响, 第 8～12 天的入水流量发生了较大变化, 同时入水的各种污染物浓度也发生了较大的波动.

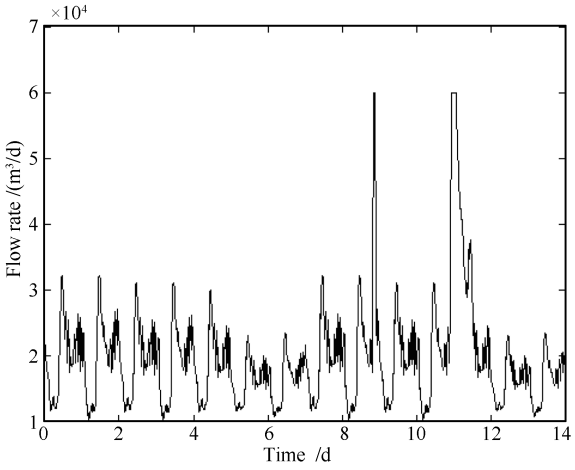

图 3　暴雨情况下的入水流量变化

Fig. 3　Storm weather influent

3.2　控制器参数选择

对于污水处理过程而言, 控制精度与出水水质直接相关, 而出水水质达标是控制的首要目标. 所以设定立即回报为

$$\boldsymbol{r}(k) = \left[\frac{1}{2}E_1^2(k), \frac{1}{2}E_2^2(k)\right]^{\mathrm{T}} \tag{41}$$

其中, $E_1(k)=y_1(k)-R_1(k)$, $E_2(k)=y_2(k)-R_2(k)$, $y_1(k)$ 和 $y_2(k)$ 分别是 $S_{\mathrm{O},5}$ 和 $S_{NO,2}$ 的测量值, $R_1(k)$ 和 $R_2(k)$ 为相应的设定值. 评价网络的输入为 $\boldsymbol{x}(k)=[y_1(k), y_2(k)]^{\mathrm{T}}$; ID ESN 的输入为 $[K_La_5(k),\ Q_a(k),\ y_1(k),\ y_2(k)]^{\mathrm{T}}$; 控制网络的输入为 $[K_La_5(k), Q_a(k), y_1(k), y_2(k), X_{BH}(k), X_{BA}(k), Q_0(k)]^{\mathrm{T}}$, $X_{BH}(k)$, $X_{BA}(k)$ 和 Q_0 分别为异养菌浓度, 自养菌浓度以及入水流量. 系统的采样周期 $T=1.25\times10^{-2}$ h $\approx$ 45 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[12] 约为 0.68, 0.68 和 0.76; 折扣因子 $\gamma=0.95$.

3.3　实验结果及分析

分别采用 PID, E-SDHP (采用 ESN 网络作为评价、控制及辨识网络的逼近器, 但回报值 Q 为标量) 及 E-MDHP 策略对系统进行控制, 三种控制器作用下, $S_{\mathrm{O},5}$ 和 $S_{\mathrm{NO},2}$ 的变化曲线如图 4 所示.

由图 4 可以看出, E-MDHP 控制器作用下, $S_{\mathrm{O},5}$ 和 $S_{\mathrm{NO},2}$ 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), E-SDHP 中 $S_{\mathrm{O},5}$ 和 $S_{\mathrm{NO},2}$ 控制性能明显下降, 而 E-MDHP 中控制性能只出现了较小的波动. 这说明 E-MDHP 具有更强的学习能力.

表 1 列出了三种控制器的控制精度指标. 可以看出, 对于溶解氧浓度, E-MDHP 的平均绝对误差 (Mean absolute error, MAE) 较 E-SDHP 及 PID 分别降低了 10 倍和 3 倍, 波动方差降低了一个数量级, 最大偏离量 $\mathrm{DEV}^{\max}$ 也明显减小. 在硝态氮浓度控制方面, E-MDHP 和 PID 的 MAE 及 $\mathrm{DEV}^{\max}$ 指标相当, 但波动方差较 PID 控制降低了一个数量级. 相比之下, E-SDHP 对硝态氮的浓度控制效果较差.

表 1 三种控制器的控制精度评价指标

Table 1 Control precision indices of three controllers

控制变量	PID		E-SDHP		E-MDHP	
	$S_{\mathrm{O},5}$	$S_{\mathrm{NO},2}$	$S_{\mathrm{O},5}$	$S_{\mathrm{NO},2}$	$S_{\mathrm{O},5}$	$S_{\mathrm{NO},2}$
MAE (绝对平均误差)	0.0442	0.0256	0.0151	0.0714	0.0041	0.0271
$\mathrm{DEV}^{\max}$ (最大偏差)	0.0993	0.1292	0.0824	0.3210	0.0145	0.1897
MSE (均方差)	4.98×10^{-4}	0.0011	3.04×10^{-4}	0.0084	2.13×10^{-5}	4.97×10^{-4}

表 2 三种控制器的控制量评价指标

Table 2 Operating amount indices of three controllers

控制变量	PID		E-SDHP		E-MDHP	
	K_La_5	Q_a	K_La_5	Q_a	K_La_5	Q_a
$\Delta u_{\max}$ (控制量最大增量)	115.01	15 674	114.13	17 549	114.29	32 675
MSE (均方差)	2 998.1	2.9875×10^{8}	2 973.5	2.1436×10^{8}	2 996.2	3.9436×10^{8}

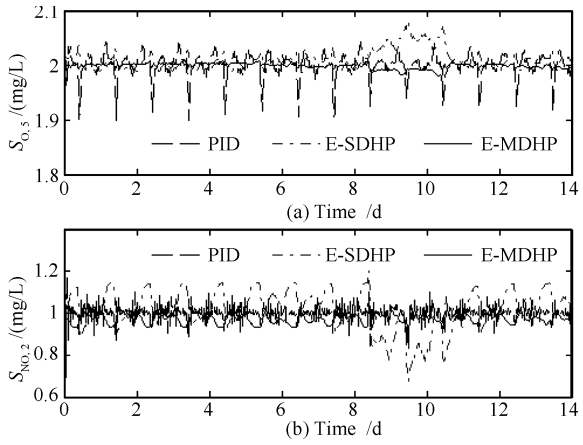

图 4 $S_{\mathrm{O},5}$ 及 $S_{\mathrm{NO},2}$ 控制效果

Fig. 4 Control effects of $S_{\mathrm{O},5}$ and $S_{\mathrm{NO},2}$

表 2 列出了三种控制器的控制量波动指标. E-MDHP 控制下, 曝气量的波动均方差 (Mean square error, MSE) 略大于 PID 及 E-SDHP, 而 Q_a 的波动方差及最大增量均明显高于 PID 及 E-SDHP. 这表明 E-MDHP 能够根据控制需求对系统状态变化做出更快速的反应, 其通过增大控制量的变化提高控制精度. 综合考虑溶解氧和硝态氮精度指标, E-MDHP 控制器明显优于 PID 及 E-SDHP 控制器.

在实验中, E-SDHP 的性能指标为 E-MDHP 的两个性能指标之和, 即:

$$Q(k)=\boldsymbol{Q}_1(k)+\boldsymbol{Q}_2(k) \tag{42}$$

$\boldsymbol{E}^c$ (式 (11)) 能够反映评价函数逼近的精确性. 对于 E-SDHP

$$\begin{cases}\boldsymbol{E}_1^c(k)=\dfrac{\partial Q}{\partial \boldsymbol{x}_1}-\gamma\dfrac{\partial Q}{\partial \boldsymbol{x}_1}-\dfrac{\partial r}{\partial \boldsymbol{x}_1}\\ \boldsymbol{E}_2^c(k)=\dfrac{\partial Q}{\partial \boldsymbol{x}_2}-\gamma\dfrac{\partial Q}{\partial \boldsymbol{x}_2}-\dfrac{\partial r}{\partial \boldsymbol{x}_2}\end{cases} \tag{43}$$

对于 E-MDHP

$$\begin{cases}\boldsymbol{E}_{11}^c(k)=\dfrac{\partial \boldsymbol{Q}_1}{\partial \boldsymbol{x}_1}-\gamma\dfrac{\partial \boldsymbol{Q}_1}{\partial \boldsymbol{x}_1}-\dfrac{\partial \boldsymbol{r}_1}{\partial \boldsymbol{x}_1}\\ \boldsymbol{E}_{21}^c(k)=\dfrac{\partial \boldsymbol{Q}_2}{\partial \boldsymbol{x}_1}-\gamma\dfrac{\partial \boldsymbol{Q}_2}{\partial \boldsymbol{x}_1}-\dfrac{\partial \boldsymbol{r}_2}{\partial \boldsymbol{x}_1}\\ \boldsymbol{E}_{12}^c(k)=\dfrac{\partial \boldsymbol{Q}_1}{\partial \boldsymbol{x}_2}-\gamma\dfrac{\partial \boldsymbol{Q}_1}{\partial \boldsymbol{x}_2}-\dfrac{\partial \boldsymbol{r}_1}{\partial \boldsymbol{x}_2}\\ \boldsymbol{E}_{22}^c(k)=\dfrac{\partial \boldsymbol{Q}_2}{\partial \boldsymbol{x}_2}-\gamma\dfrac{\partial \boldsymbol{Q}_2}{\partial \boldsymbol{x}_2}-\dfrac{\partial \boldsymbol{r}_2}{\partial \boldsymbol{x}_2}\end{cases} \tag{44}$$

综合式 (42) ∼ (44) 可得:

$$\begin{cases}\boldsymbol{E}_1^c(k)=\boldsymbol{E}_{11}^c(k)+\boldsymbol{E}_{21}^c(k)\\ \boldsymbol{E}_2^c(k)=\boldsymbol{E}_{12}^c(k)+\boldsymbol{E}_{22}^c(k)\end{cases} \tag{45}$$

所以, 从理论的角度, E-MDHP 与 E-SDHP 本质上是一致的. 但是从神经网络逼近的角度, 指标分解可以降低评价网络输入–输出关系的复杂性. 图 5 显示了学习过程中 $\boldsymbol{E}^c$ 的变化情况, 可以看出, E-MDHP 策略下 $\boldsymbol{E}^c$ 更接近于 0, 这说明 E-MDHP 策略下评价值的逼近结果更为准确.

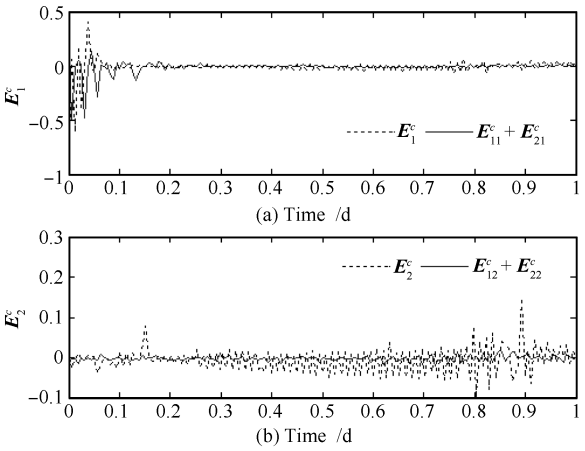

图 5 评价网络训练过程中 $\boldsymbol{E}^c$ 的变化

Fig. 5 Changes of $\boldsymbol{E}^c$ in the learning process of the critic ESN

由于内回流流量对溶解氧浓度影响极小, 所以, 在每次迭代中, 可令 $\frac{\partial \boldsymbol{Q}_1(k+1)}{\partial \boldsymbol{u}_2(k)} = 0$, 这可以消除评价指标 $\boldsymbol{Q}_1$ 对控制量 $\boldsymbol{u}_2$ 的影响, 从而加快权值调整的速度.

4 结论

强耦合的多目标控制的难点在于各控制目标之间存在冲突, 自适应动态规划的智能搜索过程实质上是一个解决和协调冲突, 进而达到一个妥协优化的过程. E-MDHP 通过指标分解提升了策略的评价精度, 进而提高了控制器的策略搜索能力, 即其冲突协调能力得到了提高. 将单一评价指标分解为多个评价指标, 还可以更为清楚地了解各控制量对每个评价指标分量的影响, 从而更利于在控制策略的搜索过程中加入先验知识. ESN 只需调整输出权值, 即各输出对应的权值调整相互独立, 易解决有较多输出的问题, 所以 ESN 适合实现多指标的 DHP 控制策略.

References

1 Shi Xiong-Wei, Qiao Jun-Fei, Yuan Ming-Zhe. Optimal control for wastewater treatment process based on improved particle optimization algorithm. *Information and Control*, 2011, **40**(5): 698−703
(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. 信息与控制, 2011, **40**(5): 698−703)

2 Holenda B, Domokos E, Rédey Á, Fazakas J. Dissolved oxygen control of the activated sludge wastewater treatment process using model predictive control. *Computers and Chemical Engineering*, 2008, **32**(6): 1270−1278

3 Dellana S A, West D. Predictive modeling for wastewater applications: linear and nonlinear approaches. *Environmental Modelling & Software*, 2009, **24**(1): 96−106

4 Zhang H G, Cui L L, Zhang X, Luo Y H. Data-driven robust approximate optimal tracking control for unknown general nonlinear systems using adaptive dynamic programming method. *IEEE Transactions on Neural Networks*, 2011, **22**(12): 2226−2236

5 Lewis F L, Vamvoudakis K G. Reinforcement learning for partially observable dynamic processes: adaptive dynamic programming using measured output data. *IEEE Transactions on Systems, Man, and Cybernetics — Part B: Cybernetics*, 2011, **41**(1): 14−25

6 Wang F Y, Zhang H G, Liu D R. Adaptive dynamic programming: an introduction. *IEEE Computational Intelligence Magazine*, 2009, **4**(2): 39−47

7 Wei Qing-Lai, Zhang Hua-Guang, Cui Li-Li. Data-based optimal control for discrete-time zero-sum games of 2-D systems using adaptive critic designs. *Acta Automatica Sinica*, 2009, **35**(6): 682−692
(魏庆来, 张化光, 崔黎黎. 基于数据自适应评判的离散 2-D 系统零和博弈最优控制. 自动化学报, 2009, **35**(6): 682−692)

8 Wei Qing-Lai, Zhang Hua-Guang, Liu De-Rong, Zhao Yan. An optimal control scheme for a class of discrete-time nonlinear systems with time delays using adaptive dynamic programming. *Acta Automatica Sinica*, 2010, **36**(1): 121−129
(魏庆来, 张化光, 刘德荣, 赵琰. 基于自适应动态规划的一类带有时滞的离散时间非线性系统的最优控制策略. 自动化学报, 2010, **36**(1): 121−129)

9 Fu J, He H B, Zhou X M. Adaptive learning and control for MIMO system based on adaptive dynamic programming. *IEEE Transactions on Neural Networks*, 2011, **22**(7): 1133−1148

10 Zhao Dong-Bin, Liu De-Rong, Yi Jian-Qiang. An overview on the adaptive dynamic programming based urban city traffic signal optimal control. *Acta Automatica Sinica*, 2009, **35**(6): 676−681
(赵冬斌, 刘德荣, 易建强. 基于自适应动态规划的城市交通信号优化控制方法综述. 自动化学报, 2009, **35**(6): 676−681)

11 White D A, Sofge D A. *Handbook of Intelligent Control: Neural, Fuzzy, and Adaptive Approaches*. New York: Van Nostrand Reinhold Press, 1992

12 Jaeger H. The "echo state" approach to analysing and training recurrent neural networks. *GMD Report, German National Research Center for Information Technology*, 2001, **12**(8): 1−43

13 Busoniu L, Babuska R, De Schutter B. *Reinforcement Learning and Dynamic Programming Using Function Approximators*. Boca Raton: CRC Press, 2010

14 IWA Taskgroup on Benchmarking of Control Stategies for WWTPs. Benchmark simulation model No. 1 (BSM1) [Online], available: http://www.iwapublishing.com, April 2008

乔俊飞 北京工业大学电子信息与控制工程学院教授. 主要研究方向为智能控制, 神经网络分析与设计. 本文通信作者.
E-mail: isibox@sina.com
(**QIAO Jun-Fei** Professor at Beijing University of Technology. His research interest covers intelligent control, analysis and design of neural networks. Corresponding author of this paper.)

薄迎春 北京工业大学智能系统研究所博士研究生. 主要研究方向为过程控制, 智能优化控制, 神经网络分析与设计.
E-mail: boyingchun@sina.com
(**BO Ying-Chun** Ph. D. candidate at the Intelligent System Institute, Beijing University of Technology. His research interest covers process control, intelligent optimal control and design of neural networks.)

韩 广 北京工业大学智能系统研究所博士研究生. 主要研究方向为智能优化控制, 神经网络分析与设计. E-mail: neraul07@126.com
(**HAN Guang** Ph. D. candidate at the Intelligent System Institute, Beijing University of Technology. His research interest covers intelligent optimal control and design of neural networks.)