

文章编号:1671-6833(2017)05-0001-06

基于互信息和 Just-in-Time 优化的回声状态网络

张 衡,王河山

(郑州大学 电气工程学院,河南 郑州 450001)

摘 要:为了提高回声状态网络(ESN)的适应性,提出基于互信息(MI)和 Just-in-Time(JIT)的优化方法,对 ESN 的输入伸缩参数以及输出层进行优化,所得网络称为 MI-JIT-ESN. ESN 的优化方法分为两部分:一是基于网络输入与输出之间的互信息,对网络的多个输入伸缩参数进行调整;二是基于 JIT 优化的局部输出层,对 ESN 的隐层输出数据进行局部重新建模,从而提升 ESN 输出层的回归拟合精度. 将 MI-JIT-ESN 应用于青霉素补料分批发酵过程建模. 结果显示,MI-JIT 优化方法能提高模型的适应性,并优于其他比较方法.

关键词:回声状态网络;互信息;Just-in-Time;优化;建模;青霉素发酵

中图分类号:TP183 文献标志码:A doi:10.13705/j.issn.1671-6833.2017.05.018

0 引言

发酵过程广泛应用于化工、制药、食品、环境科学等行业,补料分批发酵是发酵工业代表性的操作方法<sup>[1]</sup>. 发酵过程机理复杂,又具有极强的非线性、时变性以及非确定性,同时缺少关键生物参数实时测量仪器. 发酵过程的优化与控制需要对营养物和产物等重要参数进行精确预测. 因此,数据驱动的建模方法广泛地应用于补料分批发酵过程,并取得了一定的成果. 常见的数据驱动方法包括偏最小二乘法<sup>[2-3]</sup>、支持向量回归<sup>[4-5]</sup>、神经网络等<sup>[6-7]</sup>. 递归神经网络(RNN)能记忆历史信息并能体现出动态特性,非常适用于发酵过程建模.

作为一种新型递归神经网络—回声状态网络(echo state network, ESN)<sup>[8]</sup>利用“动态储备池”代替传统神经网络的隐层,提取输入的相关特征并提供期望输出动态信息,可以有效处理非线性系统辨识<sup>[9]</sup>和混沌时间序列预测问题<sup>[10]</sup>. ESN 储备池由大量的稀疏连接神经元组成,通常有几十到几百个神经元,连接权值随机产生. 与传统递归神经网络相比,ESN 最大的优势是只需训练输出权值,简化了训练过程,解决了前者训练算法过于复杂以及记忆渐消问题<sup>[11]</sup>,因此,ESN 网络更适

用于对发酵过程中重要生物参数进行建模预测. 实际应用中,ESN 良好的适应性起了关键作用. 除了结合经验知识进行的尝试以及交叉验证法之外,众多学者还提出大量的改进方法对储备池和输出权值进行优化. 对储备池的优化包括:内部拓扑结构的改善<sup>[12]</sup>,以及对储备池关键参数(包括规模、内部连接权矩阵谱半径、输入伸缩参数等)进行优化<sup>[13]</sup>. 对输出权值的优化包括:用智能算法<sup>[14]</sup>计算权值,以及对结构连接的调整等<sup>[15]</sup>,这些方法一定程度上改善了网络性能. 为了提高 ESN 网络对多输入多输出(MIMO)系统的适应性,笔者提出基于互信息(MI)和 Just-in-Time(JIT)学习的优化方法,对 ESN 网络的输入伸缩参数以及网络输出层进行优化,并将优化所得网络 MI-JIT-ESN 应用在青霉素补料分批发酵过程建模中.

1 回声状态网络

1.1 网络结构

ESN 的网络结构通常分为 3 层:输入层、状态储备池以及输出层. 假设输入层和输出层神经元个数分别为  $r$  和  $m$ ,储备层规模即神经元个数为  $n$ . 笔者采用的 ESN 网络表达式为

$$x(k) = f(W^{res}x(k-1) + W^{in}u(k)); \tag{1}$$

$$y(k) = W^{out}x(k) + b, \tag{2}$$

式中:  $x(k)$ 、 $u(k)$ 、 $y(k)$  分别为  $k$  时刻储备池的状态变量、网络输入和网络输出;  $f(\cdot)$  为储备池激活函数, 通常为双曲正切函数;  $\mathbf{W}^{\text{res}}$  是储备池的内部连接权值, 为随机产生的、稀疏连接的高维方阵(稀疏度通常在 0.01 ~ 0.05 之间, 谱半径小于 1), 一经产生则保持不变;  $\mathbf{W}^{\text{in}}$  是输入权值矩阵, 随机产生而且保持不变;  $\mathbf{W}^{\text{out}}$  和  $\mathbf{b}$  分别为输出权值矩阵和输出偏置向量。

## 1.2 网络训练

ESN 的训练过程可以转换为线性回归问题. 依次输入  $M$  训练样本后, 从  $M_0$  步开始收集储备池状态以及对应的期望输出, 构成增广状态矩阵  $\mathbf{S}$  和期望输出矩阵  $\mathbf{D}$ ,

$$\mathbf{S} = \begin{bmatrix} \tilde{x}(M_0) \\ \tilde{x}(M_0 + 1) \\ \dots \\ \tilde{x}(M) \end{bmatrix},$$

其中:  $\tilde{x}(k) = [x(k)^T \quad 1]^T$ .

$$\mathbf{D} = \begin{bmatrix} d(M_0)^T \\ d(M_0 + 1)^T \\ \dots \\ d(M)^T \end{bmatrix}.$$

因此, 输出权值和偏置可由下式求得,

$$[\mathbf{W}^{\text{out}} \quad \mathbf{b}]^T = (\mathbf{S}^T \cdot \mathbf{S})^{-1} \cdot \mathbf{S}^T \cdot \mathbf{D}. \quad (3)$$

## 2 基于互信息和 JIT 优化的回声状态网络

### 2.1 互信息理论

互信息 (mutual information) 是信息论里一种用来体现两个事件集合之间相关性的理论概念<sup>[16]</sup>. 给定两个随机变量  $X$  和  $Y$ , 若它们各自的边缘概率分布和联合概率分布分别为  $p(x)$ 、 $p(y)$  和  $p(x, y)$ , 则它们之间的互信息  $I(X; Y)$  定义为

$$I(X; Y) = \sum_x \sum_y p(x, y) \log \frac{p(x, y)}{p(x)p(y)}. \quad (4)$$

笔者采用 Pocock 的互信息工具箱计算互信息, 其中将多维变量转化为一个联合随机变量, 更好地计算了两个事件集合之间的互信息<sup>[17]</sup>.

### 2.2 Just-in-Time 学习

JIT 学习算法是目前很流行的一种在线过程监控建模方法, JIT 学习模型是根据当前需测量数据和历史训练数据的最近空间距离来建立多个局部模型的建模方法. JIT 学习模型可以更好地处理过程的非线性特性, 从而提升整个过程的测量和预测精度.

笔者对 ESN 的输出层进行 JIT 学习优化, 由于 ESN 的输出层维数较高 (通常在 100 以上), 输出层进行回归拟合时, 处理的数据量很大, 通过 JIT 建立输出层的局部回归模型可以有效地提升 ESN 回归拟合时的精度.

### 2.3 MI-JIT-ESN 建模过程

将用于学习的数据归一化, 并分为 3 部分: 训练数据、验证数据和测试数据.

建立 MI-JIT-ESN 模型的主要步骤如下:

(1) 初始化  $r$  输入  $m$  输出的 ESN 网络. 默认输入缩放参数均为 1, 储备池规模和谱半径可由经验法或试凑法确定.

(2) 基于互信息的输入伸缩参数优化. ④计算输入与输出间互信息. 依据训练数据中的输入数据与输出数据, 计算每一个输入变量与输出之间的互信息, 得到互信息  $I = [I_1 \quad I_2 \quad \dots \quad I_r]$ . 令

$$\beta = [\beta_1 \quad \beta_2 \quad \dots \quad \beta_r] = \frac{\alpha}{I_{\max}} \cdot I, \quad (5)$$

式中:  $\alpha$  为调整系数;  $I_{\max}$  为最大互信息值. ⑤对  $\beta$  进行调整. 从 0.1 开始以 0.1 为间隔逐步增大  $\beta$ , 直到验证误差增大, 即 ESN 网络对验证数据的处理能力下降为止, 所得到的缩放参数为  $\hat{\beta}$ . ⑥保存以  $\hat{\beta}$  为输入缩放参数的新网络.

(3) 更新所得网络的状态. 收集得到的增广状态矩阵  $\mathbf{S}$  可以表示为

$$\mathbf{S} = [s_1, s_2, \dots, s_n, s_0]. \quad (6)$$

式中:  $s_0$  为均是 1 的列向量,  $s_l (l = 1, 2, \dots, n)$  为储备层第  $l$  个神经元的输出列向量. 期望输出矩阵  $\mathbf{D}$  表示为

$$\mathbf{D} = [d_1, d_2, \dots, d_m]. \quad (7)$$

式中:  $d_j (j = 1, 2, \dots, m)$  为输出层第  $j$  个神经元的输出列向量.

(4) 在网络训练阶段得到 ESN 输出层的所有状态, 即训练数据库  $\mathbf{S}_{\text{train}}$ , 然后针对每一组新的验证数据, 用距离测度方法在训练数据库  $\mathbf{S}_{\text{train}}$  中选取和当前验证样本最近距离的子训练样本集, 子训练样本集的长度为  $L$ ,  $L$  的值采用试凑法确定, 建立局部 ESN 输出回归模型.

(5) 对每个输出重复步骤 (4), 得到最终的网络结构, 即 MI-JIT-ESN 模型.

## 3 应用

### 3.1 青霉素补料分批发酵过程

采用 MI-JIT-ESN 方法对青霉素补料分批发酵过程进行建模. 本研究中, 过程数据来自基于内

核的 Pensim 发酵仿真平台<sup>[18]</sup>.

研究中,选用空气流量、搅拌功率、底物流加速率、底物温度、溶解氧浓度、排气二氧化碳浓度、培养体积、发酵罐中 pH 值、发酵罐中温度和产生的热量这 10 个可在线测量的变量作为输入变量,选用底物浓度、菌体浓度、青霉素产物浓度这 3 个无法在线测量的关键生物变量作为输出变量进行建模研究.用 Pensim 平台生成 10 个批次的数据用来建立和优化模型,另外 5 个批次的数据用来评估模型的准确性和可靠性.一个批次的反应时间为 400 h,采样时间为 0.5 h,所以每个批次有 800 个样本点.同时,利用测试数据的均方根误差 (RMSE) 来判断所建模型的性能.含有  $N$  个样本点的一个批次中,第  $j$  个输出变量

$$RMSE_j = \sqrt{\frac{1}{N} \sum_{k=1}^N (ya_j(k) - yp_j(k))^2}, \quad (8)$$

式中:  $ya_j(k)$ 、 $yp_j(k)$  分别为  $k$  时刻第  $j$  个输出的实际值和模型预测值.

3.2 青霉素发酵过程的 MI-JIT-ESN 模型

首先将默认状态下 10 个批次的学习数据分别归一化,其中 6 个批次数据用来训练,4 个批次作为验证数据.建立 10 输入、3 输出、初始储备池规模为 500 的 ESN 网络结构,并初始化,内部连接权重矩阵谱半径定为 0.9.

(1)伸缩参数的调整.由输入与输出数据互信息可得到伸缩参数  $\hat{\beta}$ ,基于 MI 信息的试凑法过程如图 1 所示.可得这次选择  $\alpha = 0.6$ ,此时 10 个输入伸缩参数为:

$$[\beta_1 \ \beta_2 \ \cdots \ \beta_{10}] = [0.04 \ 0.02 \ 0.26 \ 0.02 \ 0.25 \ 0.60 \ 0.34 \ 0.15 \ 0.06 \ 0.38]. \quad (9)$$

JIT 局部回归模型的建立.笔者的输出层训练数据长度为 4 800,维度为 500 维,在不同的子样本训练集长度下,分别建立局部回归模型,并用验证集进行测试,测试结果如表 1 所示.

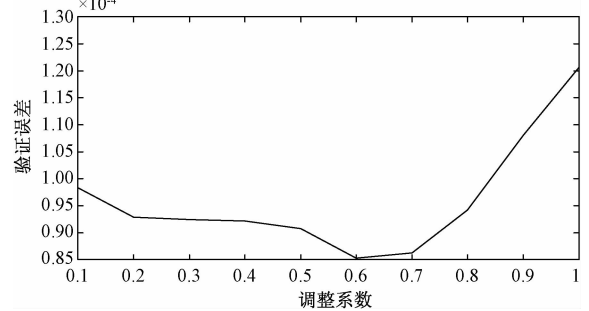


图 1 网络验证误差随调整系数  $\alpha$  变化情况  
Fig.1 Network validation error with the change of adjustment factor  $\alpha$

表 1 MI-JIT-ESN 模型在不同子样本训练集长度下的测试结果

Tab.1 The performance of MI-JIT-ESN models in different subsample length

$L$	基质浓度	生物量浓度	青霉素浓度
4 800	0.028 5	0.002 3	0.013 7
4 600	0.028 0	0.001 9	0.012 9
4 400	0.027 8	0.001 9	0.012 5
4 200	0.027 5	0.001 8	0.012 2
4 000	0.027 0	0.001 7	0.012 0
<b>3 800</b>	<b>0.026 1</b>	<b>0.001 5</b>	<b>0.011 5</b>
3 600	0.026 3	0.001 6	0.012 0
3 400	0.026 6	0.001 8	0.012 4
3 200	0.027 1	0.002 0	0.012 7
3 000	0.027 3	0.002 1	0.013 0

由表 1 可知,在子训练集样本长度为 3 800 时,验证数据的测试效果最好.最终  $L$  取 3 800 建立局部回归模型,然后再对网络进行测试.

3.3 结果与对比

(1)采用未参与建模优化过程的全新测试数据来检验所得 MI-JIT-ESN 模型的效果.

优化后的 MI-JIT-ESN 模型与未优化的 ESN 模型对 3 个关键生物量:底物浓度、菌体浓度、青霉素浓度的一个批次内预测效果如图 2 ~ 4 所示.可以看出 MI-JIT-ESN 模型很好地同时预测了青霉素补料分批发酵过程中的 3 个关键生物量.为了更好地显示优化效果,图 5 展示了未优化 ESN 模型和 MI-JIT-ESN 模型对一个测试批次 3 个关键生物量的预测误差之间的对比.表 2 为优化前后的模型对多个不同批次的测试 RMSE 值.与未优化 ESN 模型相比,笔者所提出的 MI-JIT-ESN 模型大大提高了对青霉素发酵过程中生物量的预测效果.

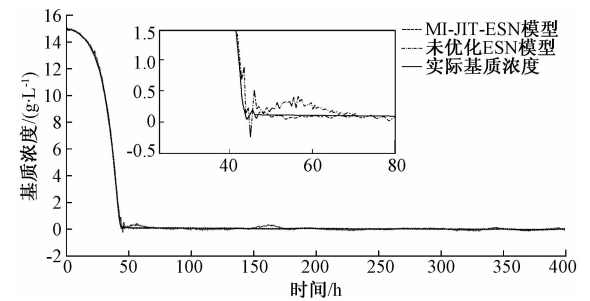


图 2 采用未优化 ESN 和 MI-JIT-ESN 方法对一个测试批次基质浓度的预测效果对比  
Fig.2 Trend plots of substrate concentration predictions at one testing batch using unoptimized ESN and MI-JIT-ESN methods

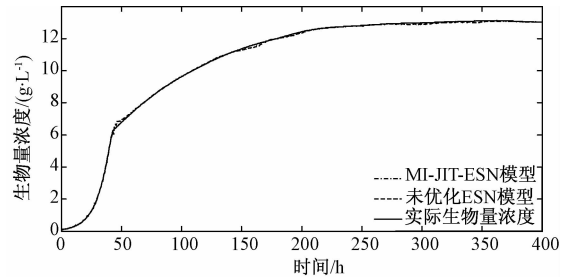


图 3 采用未优化 ESN 和 MI-JIT-ESN 方法对一个测试批次生物量浓度的预测效果对比

Fig.3 Trend plots of biomass concentration predictions at one testing batch using unoptimized ESN and MI-JIT-ESN methods

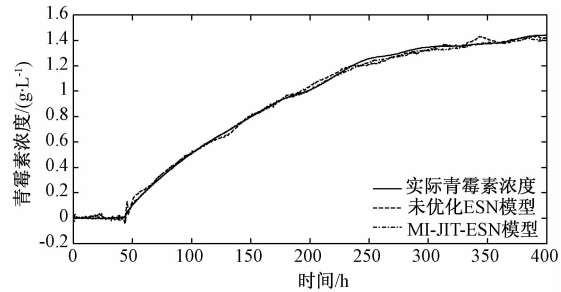


图 4 采用未优化 ESN 和 MI-ESN 方法对一个测试批次青霉素浓度的预测效果对比

Fig.4 Trend plots of penicillin concentration predictions at one testing batch using unoptimized ESN and MI-JIT-ESN methods

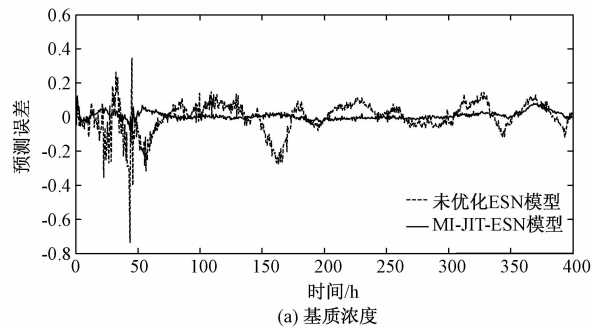
(2)在不同初始底物和菌体浓度条件下,青霉素发酵过程反应很不同.为进一步检验模型的性能,对低工况、中工况、高工况进行了仿真研究,具体设定条件如表 3 所示(其余条件采用默认值).

表 2 MI-JIT-ESN 模型与未优化 ESN 模型的测试 RMSE 值对比

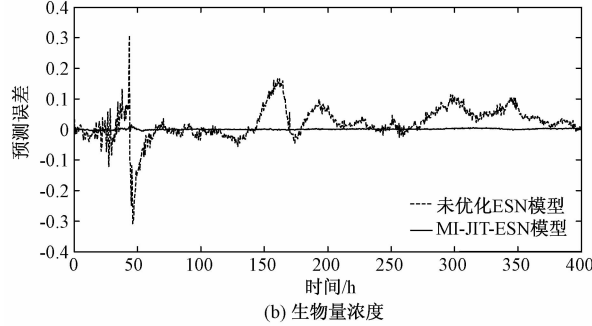
Tab.2 Comparison of testing RMSE values using unoptimized ESN and MI-JIT-ESN methods

编号	基质浓度		生物量浓度		青霉素浓度	
	ESN	MI-JIT-ESN	ESN	MI-JIT-ESN	ESN	MI-JIT-ESN
1	0.152 7	0.040 9	0.053 1	0.001 9	0.026 0	0.010 2
2	0.087 6	0.017 5	0.049 5	0.001 3	0.014 9	0.009 8
3	0.091 3	0.021 4	0.056 4	0.001 8	0.022 7	0.013 3
4	0.114 6	0.021 1	0.054 7	0.001 1	0.020 4	0.010 4
5	0.172 2	0.030 8	0.080 0	0.001 7	0.021 1	0.015 7
平均值	0.123 9	0.026 3	0.058 8	0.001 6	0.021 0	0.011 9

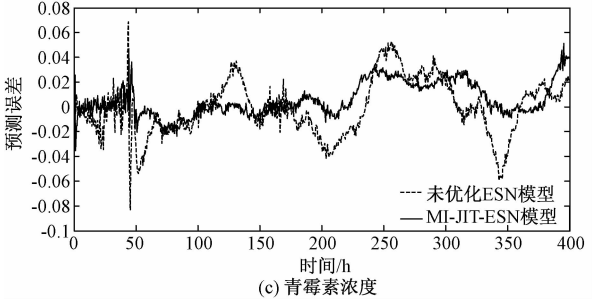
将 MI-JIT-ESN 模型与传统递归神经网络 Elman 网络模型以及前向网络 BP 网络模型作对比,Elman 神经网络和 BP 神经网络的隐层神经元个数都选 15 个,训练算法都采用传统的 LM 算法.比较标准为 5 个测试批次的平均 RMSE 误差 MRMSE,如表 4 所示.表 4 所列结果为多次试验平均值,可以看出,大部分情况下,MI-JIT-ESN 模型优于



(a) 基质浓度



(b) 生物量浓度



(c) 青霉素浓度

图 5 采用未优化 ESN 和 MI-JIT-ESN 方法对一个测试批次 3 个关键生物量的预测误差对比

Fig.5 Prediction errors of three key biological variables at one testing batch using unoptimized ESN and MI-JIT-ESN methods

表 3 不同工况的初始条件

Tab.3 The initial conditions of different working conditions

初始条件	基质浓度/ ( $\text{g}\cdot\text{L}^{-1}$ )	生物量浓度/ ( $\text{g}\cdot\text{L}^{-1}$ )	基质进给率/ ( $\text{L}\cdot\text{h}^{-1}$ )
低工况	4.5 ~ 5.5	0.004 ~ 0.005	0.035
中工况	14 ~ 16	0.09 ~ 0.12	0.045
高工况	23 ~ 27	0.18 ~ 0.22	0.070

传统神经网络模型,尤其在菌体浓度的预测和高工况即初始底物和菌体浓度较大的情况下.

表 4 不同神经网络模型的测试 MRMSE 值对比  
Tab.4 Comparison of testing MRMSE values using different neural network methods

初始条件	NN 模型	基质 浓度	生物量 浓度	青霉素 浓度
低工况	MI-JIT-ESN	0.033 1	0.002 4	0.012 5
	Elman	0.020 2	0.032 5	0.012 7
	BP	0.047 7	0.660 1	0.089 3
中工况	MI-JIT-ESN	0.030 1	0.001 8	0.013 0
	Elman	0.043 9	0.064 4	0.015 4
	BP	1.229 7	0.771 8	0.110 7
高工况	MI-JIET-ESN	0.029 0	0.003 3	0.018 5
	Elman	0.070 1	0.069 9	0.021 3
	BP	2.104 0	1.112 6	0.156 2

为了验证 MI-JIT-ESN 模型的性能,笔者选取了一些最新文献中的青霉素发酵过程预测模型,例如:Relevance Vector Machine (RVM)<sup>[19]</sup>、BI-SVR<sup>[20]</sup>、核偏最小二乘法(KPLS)和 MGMM<sup>[21]</sup>,并与其作比较,比较结果如表 5.

表 5 正常工况下 MI-JIT-ESN 模型和不同青霉素发酵模型的测试 MRMSE 值对比  
Tab.5 Comparison of MRMSE values of quality variable predictions using MI-JIT-ESN and other methods on normal conditions

方法	基质浓度	生物量浓度	青霉素浓度
RVM	—	0.192 7	0.042 8
KPLS	0.541 0	0.897 0	0.195 0
MGMM	0.073 0	0.092 0	0.038 0
BI-SVR	0.085 0	0.071 0	0.011 0
MI-JIT-ESN	0.020 9	0.001 5	0.012 0

从表 5 中我们可以看出,MI-JIT-ESN 在青霉素建模过程中性能良好,因此,笔者所提的 MI-JIET-ESN建模方法在青霉素批处理建模过程中是一个更优的选择,同时,笔者所提的模型也适合和青霉素批处理建模相似的建模过程.

4 结论

笔者提出一种基于互信息和 JIT 学习的优化方法对 ESN 进行优化,以提高其对 MIMO 系统的适应性.该方法中,不仅基于网络的输入与输出之间的互信息对网络的多个输入伸缩参数进行调整优化;而且基于 JIT 学习,对 ESN 输出层回归部分进行局部重新建模,达到结构优化的目的.将该方

法运用到对青霉素补料分批发酵过程中关键生物参数的建模中,与优化前的 ESN 模型的测试效果比较,取得良好的优化效果,达到了提高适应度的目的.另外,在不同情况下的建模结果体现出了 MI-JIT-ESN 模型的有效性与准确性,而与其他传统神经网络建模方法的对比也体现出了 MI-JIT-ESN 模型的优越性.

参考文献:

[1] LEE J, LEE S Y, PARK S, et al. Control of fed-batch fermentations [J]. Biotechnology advances, 1999, 17(1):29-48.

[2] 周彩荣,刘中平. 偏最小二乘-速差动力学分光光度法同时测定  $S_n^{2+}$  与  $S_n^{4+}$  [J]. 郑州大学学报(理学版), 2015, 47(1):97-102.

[3] 李宗坤,陈乐意,孙颖章. 偏最小二乘回归在渗流监控模型中的应用[J]. 郑州大学学报(工学版), 2006, 27(2):117-119.

[4] 杨高飞,徐睿,秦鸣,等. 基于 ARMA 和卡尔滤波的短时交通预测[J]. 郑州大学学报(工学版),2017, 38(2):36-40.

[5] 李凌均,陈超,韩捷,等. 全矢支持向量回归频谱预测方法[J]. 郑州大学学报(工学版), 2016, 37(3):78-82.

[6] 胡燕,宋晓瑛,马刚. 基于 K-means 和时间匹配的预测模型[J]. 郑州大学学报(工学版), 2017, 38(2):17-2 .

[7] 王杰,陈春宇. 基于神经网络逆系统的循环流化床内模控制[J]. 郑州大学学报(工学版), 2013, 34(2):71-75.

[8] JAEGER H. The “echo state” approach to analysing and training recurrent neural networks-with an erratum note[R]. Bonn, Germany: German National Research Center for Information Technology GMD Technical Report, 2001(148):34.

[9] JAEGER H. Adaptive nonlinear system identification with echo state networks//Advances in neural information processing systems[C]. 2002: 593-600.

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

[11] PENG Y, WANG J, PENG X. Survey on reservoir computing[J]. Acta electronica Sinica, 2011, 39(10):2387-96.

[12] RODAN A, TINO P. Minimum complexity echo state network[J]. IEEE Transactions on neural networks, 2011, 22(1):131-44.

- [13] VERSTRAETEN D, SCHRAUWEN B, Haene M, et al. An experimental unification of reservoir computing methods[J]. *Neural networks*, 2007, 20(3):391–403.
- [14] WANG J S, HAN S, GUO Q P. Echo state networks based predictive model of vinyl chloride monomer convention velocity optimized by artificial fish swarm algorithm[J]. *Soft computing*, 2014, 18(3):457–68.
- [15] DUTOIT X, SCHRAUWEN B, VAN C J, et al. Pruning and regularization in reservoir computing[J]. *Neurocomputing*, 2009, 72(7):1534–46.
- [16] GUO W. Structural optimization algorithm for rbf neural network based on mutual information[J]. *Computer science*, 2013, 40(6):252–5.
- [17] BROWN G, POCKOCK A, ZHAO M. Conditional likelihood maximisation: a unifying framework for information theoretic feature selection[J]. *The journal of machine learning research*, 2012, 13(1):27–66.
- [18] BIROL G, ÜNDEY C, Cinar A. A modular simulation package for fed-batch fermentation: penicillin production[J]. *Computers & chemical engineering*, 2002, 26(11):1553–65.
- [19] LIU N, ZHANG J, YAN J. Weighted kernel regression for semi-supervised soft-sensing modeling of fed-batch processes[J]. *Taiwan Inst. Chem. Eng.* 2012(43):67–76.
- [20] YU J. A Bayesian inference based two-stage support vector regression framework for soft sensor development in batch bioprocesses[J]. *Computers & chemical engineering*, 2012(41):134–144.
- [21] YU J. Multiway gaussian mixture model based adaptive kernel partial least squares regression method for soft sensor estimation and reliable quality prediction of non-linear multiphase batch processes[J]. *Industrial & engineering chemistry research*, 2012(51):13227–13237.

## Optimized Echo State Network on the Basis of Mutual Information and Just-in-time

ZHANG Heng, WANG Heshan

(School of Electrical Engineering, Zhengzhou University, Zhengzhou 450001, China)

**Abstract:** To improve the adaptability of echo state network (ESN), an optimization method based on mutual information (MI) and Just-In-Time (JIT) learning was proposed in this paper to optimize the input scaling and the output layer of ESN. The method was named as MI-JIT optimization method and the obtained new network was MI-JIT-ESN. The optimization method mainly consists of two parts. Firstly, the scaling parameters of multiple inputs were adjusted on the basis of MI between the network inputs and outputs. Secondly, based on JIT learning, a partial model of output layer was established. The new partial model could make the regression results more accurate. Further, a multi-input multi-output MI-JIT-ESN model was developed for the fed-batch penicillin fermentation process. The experimental results showed that the obtained MI-JIT-ESN model performed well, and that it had better adaptability than ESN model without optimization and other neural network models.

**Key words:** echo state network; mutual information; Just-in-Time learning; optimization; model; penicillin fermentation