

# 基于熵准则的鲁棒的 RBF 谷胱甘肽发酵建模

谭左平<sup>1,2</sup>, 王士同<sup>1</sup>, 邓赵红<sup>1</sup>, 堵国成<sup>2</sup>

江南大学信息工程学院, 江南大学工业技术教育部重点实验室, 无锡 214122

**摘要:** 在谷胱甘肽的发酵过程建模中, 当试验数据含有噪音时, 往往会导致模型预测精度和泛化能力的下降。针对该问题, 提出了一种新的基于熵准则的 RBF 神经网络建模方法。与传统的基于 MSE 准则函数的建模方法相比, 新方法能从训练样本的整体分布结构来进行模型参数学习, 有效地避免了传统的基于 MSE 准则的 RBF 网络的过学习和泛化能力差的缺陷。将该模型应用到实际的谷胱甘肽发酵过程建模中, 实验结果表明: 该方法具有较高的预测精度、泛化能力和良好的鲁棒性, 从而对谷胱甘肽的发酵建模有潜在的应用价值。

**关键词:** 谷胱甘肽, 相对熵, RBF 神经网络, Parzen 窗法, 鲁棒性

## GSH Fermentation Process Modeling Using Entropy-criterion Based RBF Neural Network Model

Zuoping Tan<sup>1,2</sup>, Shitong Wang<sup>1</sup>, Zhaohong Deng<sup>1</sup>, and Guocheng Du<sup>2</sup>

Key Laboratory of Industrial Biotechnology, Ministry of Education, School of Information Engineering, Jiangnan University, Wuxi 214122, China

**Abstract:** The prediction accuracy and generalization of GSH fermentation process modeling are often deteriorated by noise existing in the corresponding experimental data. In order to avoid this problem, we present a novel RBF neural network modeling approach based on entropy criterion. It considers the whole distribution structure of the training data set in the parameter learning process compared with the traditional MSE-criterion based parameter learning, and thus effectively avoids the weak generalization and over-learning. Then the proposed approach is applied to the GSH fermentation process modeling. Our results demonstrate that this proposed method has better prediction accuracy, generalization and robustness such that it offers a potential application merit for the GSH fermentation process modeling.

**Keywords:** GSH, relative entropy, RBF neural network, Parzen window, robustness

谷胱甘肽 (GSH) 是由 L-谷氨酸、L-半胱氨酸和甘氨酸缩合而成的一种同时含有  $\gamma$ -谷氨酰基和巯基的生物活性三肽。它可以迅速增强机体的免疫力, 同时对于维持生物体内适宜的氧化还原环境起着至

关重要的作用<sup>[1,2]</sup>。作为一种重要的生理活性物质, 谷胱甘肽在解毒、抗辐射、肿瘤、癌症、氧化衰老和协调内分泌的治疗中效果明显且无副作用<sup>[3]</sup>, 近年来还发现 GSH 具有抑制艾滋病病毒的功效<sup>[4]</sup>。此

**Received:** October 8, 2007; **Accepted:** January 16, 2008

**Supported by:** the National 863 project of China (No. 2007AA1Z158, 2006AA10Z313), National Science Foundation of China (No. 60773206/F020106, 60704047/F030304), New century Outstanding Young Scholar Grant of Ministry of Education of China (NCET-04-0496), 2006 Outstanding Young Scholar Grant at Jiangsu Province, Research Grant from National KeySoft Laboratory at Nanjing University.

**Corresponding author:** Zuoping Tan. Tel: +86-13861742694; Email: zuoping\_tantan@126.com

国家 863 资助项目 (No. 2007AA1Z158, 2006AA10Z313); 国家自然科学基金资助项目 (No. 60773206/F020106, 60704047/F030304); 2004 年教育部跨世纪优秀人才支持计划基金项目 (NO. NCET-04-0496), 2006 年江苏省 6 大人才高峰计划资助项目, 南京大学软件新技术国家重点实验室开放课题。

外, GSH 作为一种多功能的生物活性添加剂在食品加工业中的应用也会愈来愈广。随着 GSH 的生理生化功能和性质被不断研究发现, 人们对其在医药工业、食品工业、体育运动领域及有关生物研究领域上的兴趣将日益增长, 对其需求量也将不断增加。很显然, GSH 有一个极其巨大的市场。

目前, 微生物发酵法是 GSH 工业化生产的最普遍方法之一, 而且生产 GSH 所用的酵母大都是工业常用菌株, 能在很宽的培养条件下生长并积累 GSH, 但产量往往很低。因此, 要想进一步提高 GSH 的合成能力, 必须对培养基中的各种营养成分以及培养条件进行优化, 在最佳的工艺条件下进行生产, 为最终的工业化生产奠定基础。20 世纪九十年代起, 神经网络模型便开始被运用于发酵过程参数的预测及估计<sup>[5]</sup>, 并且在实际应用中, 对培养基的设计表现出较高的预测精度<sup>[6]</sup>, 是一种有较大发展前途和使用价值的模型方法。Desai 等<sup>[7]</sup>将神经网络应用于乳酸杆菌发酵生成多糖(EPS)的过程中, 张光亚等<sup>[8]</sup>研究了木聚糖酶氨基酸组成与其最适 pH 的神经网络模型, 林毅等将支持向量机模型运用于苏云金杆菌杀虫晶体蛋白活性的预测<sup>[9]</sup>和几丁质酶的最适 pH 建模<sup>[10]</sup>, 王健等<sup>[11]</sup>利用 BP 神经网络对 L-色氨酸的发酵过程进行了建模。但是, 该类技术存在的一个缺点是: 它们普遍采用的传统的基于 MSE 准则的误差函数, 即以样本的实际输出和神经网络的预测输出的平方差为误差准则函数来进行参数学习。由于 MSE 准则是基于局部逼近原理, 它只考虑了各个样本点处的逼近情况, 要求所有样本点的经验误差达到最小, 而没有考虑样本之间的空间, 这使得在某些情况下出现过学习现象而降低了模型的泛化能力和鲁棒性。

针对该问题, 本文借助于信息论中关于熵的概念, 从概率密度的角度出发<sup>[12]</sup>, 利用判别熵构造出一个新的误差准则函数, 并将其成功地应用于谷胱甘肽的发酵建模过程中。同传统的基于 MSE 准则的误差函数相比, 本文提出的基于熵准则的误差函数从所有样本整体出发, 故暗含了样本潜在的分布结构, 因此能够从全局的角度对神经网络的参数进行学习, 从而有效地增强了学习得到的模型的泛化能力和鲁棒性。我们的实验结果亦证实了上述优点。

## 1 材料与方法

### 1.1 菌株

产朊假丝酵母(*Candida utilis*)WSH 02-08, 江南大学环境生物技术研究室保藏。

### 1.2 培养基

斜面培养基(g/L): 葡萄糖 20, 蛋白胨 20, 酵母膏 10, 琼脂 20, pH 6.0。

种子培养基(g/L): 葡萄糖 20, 蛋白胨 20, 酵母膏 10, pH 6.0。

发酵基本培养基(g/L): 葡萄糖 30, 硫酸铵 8, 磷酸二氢钾 2.5, 硫酸镁 0.3, pH 5.5。各营养成分及其浓度还需要进一步考察。

### 1.3 培养方法

斜面培养: 接种后的斜面置于 30°C 恒温培养箱中培养 3 d, 冰箱冷藏室中保藏, 每个月至少需要转接一次。

种子培养: 将斜面种子在 30°C 活化 3~4 h 后, 取一环酵母菌体接种至装有 50 mL 种子培养基的 500 mL 三角瓶中培养, 摇床转速 200 r/min, 温度 30°C, 培养时间 20 h。

发酵培养: 将培养好的种子培养液按照 10%(V/V)的接种量, 接种至装有 50 mL 发酵培养基的 500 mL 三角瓶中进行发酵培养, 发酵时间 26 h, 发酵温度 30°C, HYG-II 型回转式恒温调速摇瓶柜的转速为 200 r/min。每批实验设计都安排有 3 组平行对照样, 取实验结果的平均值进行分析和计算。

### 1.4 测定方法

细胞干重(Dry Cell Weight, DCW)测定: 取 25 mL 发酵液, 经 3000 r/min 下离心后再用蒸馏水洗涤 2 次, 得到的湿酵母细胞在 60°C 下烘至恒重。

葡萄糖浓度的测定: 3,5-二硝基水杨酸法测定<sup>[13]</sup>。

胞内 GSH 的提取: 新鲜酵母经蒸馏水洗涤 3 次后, 30°C 下, 40%(V/V)的稀乙醇溶液中振荡处理 2 h, 3000 r/min 离心得到的上清液中富含 GSH, 经稀释后用作 GSH 的分析检测。

谷胱甘肽测定: DTNB [5,5' -二硫双-(2-硝基苯甲酸)]-GSH 还原酶循环法<sup>[14]</sup>。

胞内 GSH 含量(GSH content)的定义: GSH 浓度对细胞干重的百分比。

## 1.5 正交优化试验

根据 *C. utilis* WSH 02-08 进行 GSH 生产时的营养需求，确定了营养条件正交试验的初始条件如表 1 所示，采用  $L_{16}(4^5)$  正交试验表安排正交优化试验，试验数据如表 2 所示。

## 2 径向基函数(RBF)神经网络模型

### 2.1 基于 MSE 准则的 RBF 神经网络模型

在目前的建模方法中，径向基函数(RBF)神经网络的回归建模是有效的方法之一。同 BP 神经网络一样，径向基函数神经网络也是一种前馈型神经网络，然而该网络在逼近能力、分类能力和学习速度等方面均优于目前广泛采用的 BP 网络。这种网络的结构特点是：只有一个隐层，输入层节点只传递输入信号到隐层，隐层单元采用径向基函数作为其输出特性，输出层节点则通常是简单的线性求和单元。RBF 回归建模过程一般分为两个步骤：第一步是构造初始化网络模型，常用的方法有平均分割法，聚类法<sup>[15,16]</sup>；第二步是对初始化的模型参数进行学习，广泛采用的方法是基于 MSE 准则的模型。即以样本的实际输出和神经网络的预测输出的平方差作为误差准则函数(MSE 准则函数)：

$$E_1 = \frac{1}{2N} \sum_{i=1}^N (y_i - y_{di})^2 \quad (1)$$

从(1)式可以看出，传统的建模过程是一个局部逼近过程，即点点逼近。它从单个样本的角度寻找误差函数的最小，仅仅只考虑了各个样本点处的逼近情况，而忽略了样本之间的空间。由于该准则要求所有样本点的整体均方误差达到最小值，因此学习所得的网络的泛化能力和鲁棒性得不到保证，可能会出现过拟合问题。特别是当试验数据存在噪音的情况下，学习得到的模型性能会恶化。下面将从一个新的角度进行探讨，即从熵的角度提出了一种新的 RBF 神经网络建模方法。

### 2.2 基于熵准则的 RBF 神经网络模型

在信息论中用“熵”作为不确定性的一种度量的表达式，它是概率密度分布的函数。这样一个概念也可以用来作为某个概率密度分布  $f_1(x_i)$  偏离给定的标准分布  $f_2(x_i)$  的程度的度量，此时被称之为相对熵。即：

$$V(f_1, f_2) = -\sum f_1(x_i) \log[f_1(x_i)/f_2(x_i)] \leq 0 \quad (2)$$

表 1  $L_{16}(4^5)$  正交试验初始条件

Table 1 Initial condition for  $L_{16}(4^5)$  orthogonal experiments

Factors	Levels			
	1	2	3	4
A: Glucose /(g/L)	25	30	35	40
B: $(\text{NH}_4)_2\text{SO}_4$ /(g/L)	2	3	4	5
C: Urea /(g/L)	3	4	5	6
D: $\text{KH}_2\text{PO}_4$ /(g/L)	1.5	2	2.5	3
E: $\text{MgSO}_4$ /(g/L)	0.2	0.25	0.3	0.35

表 2  $L_{16}(4^5)$  正交试验数据表

Table 2 Data of the orthogonal experiments

No.	Factors and levels					Results of experiments		
	A	B	C	D	E	GSH	DCW	Intracellular GSH
1	1	1	1	1	1	166.0	6.9	2.40
2	1	2	2	2	2	163.5	7.0	2.32
3	1	3	3	3	3	167.0	6.8	2.44
4	1	4	4	4	4	160.7	6.8	2.38
5	2	1	2	3	4	158.2	7.7	2.06
6	2	2	1	4	3	151.0	7.6	1.98
7	2	3	4	1	2	156.6	7.6	2.06
8	2	4	3	2	1	152.6	7.6	2.02
9	3	1	3	4	2	151.3	7.8	1.95
10	3	2	4	3	1	142.5	7.5	1.91
11	3	3	1	2	4	143.5	7.9	1.81
12	3	4	2	1	3	144.7	8.2	1.77
13	4	1	4	2	3	143.5	7.4	1.95
14	4	2	3	1	4	144.7	7.8	1.85
15	4	3	2	4	1	151.0	8.4	1.80
16	4	4	1	3	2	144.7	8.2	1.76

由(2)式可得：相对熵越小，则这两类概率密度分布的差别就越大，反之，相对熵越大，这两类概率密度分布的差别就越小，当两类概率分布完全相同时，相对熵达到最大值(等于零)。因此可用如下的判别熵  $W(f_1, f_2)$  来表征两类分布  $f_1(x_i)$  和  $f_2(x_i)$  的差别大小。

$$\begin{aligned} W(f_1, f_2) &= V(f_1, f_2) + V(f_2, f_1) \\ &= -\sum f_1(x_i) \log[f_1(x_i)/f_2(x_i)] - \sum f_2(x_i) \log[f_2(x_i)/f_1(x_i)] \\ &= -\sum f_1(x_i) \log[f_1(x_i)] - \sum f_2(x_i) \log[f_2(x_i)] + \sum f_1(x_i) \log[f_2(x_i)] + \sum f_2(x_i) \log[f_1(x_i)] \end{aligned} \quad (3)$$

给定训练样本集  $\{x_i, y_{di}\}$ ,  $x_i \in R^d$ ,  $y_{di} \in R$ ,  $i=1, 2, \dots, N$  对于 RBF 神经网络，我们可以构造出两个新

的数据集合, 即样本输入  $\mathbf{x}_i$  和样本输出  $y_{di}$  构成的向量集合  $\mathbf{z}'_i \in R^d$  (其中  $d=d+1$ ) 以及样本输入  $x_i$  和神经网络的预测输出  $y_i$  构成的向量集合  $\mathbf{z}''_i \in R^d$ , 即:  $S_1 = \{\mathbf{z}'_i | \mathbf{z}'_i = (\mathbf{x}_i, y_{di})\}$ ,  $S_2 = \{\mathbf{z}''_i | \mathbf{z}''_i = (\mathbf{x}_i, y_i)\}$ 。

根据统计学中的相关原理, 给定样本集合  $\{\mathbf{x}_i\}$ ,  $\mathbf{x}_i \in R^d$ , 利用 Parzen 窗法<sup>[12]</sup>可以估计出其样本空间中任意一点的概率密度函数:

$$f(\mathbf{x}, \sigma) = \frac{1}{N} \sum_{i=1}^N (\sqrt{2\pi}\sigma)^{-d} e^{-\frac{\|\mathbf{x}-\mathbf{x}_i\|^2}{2\sigma^2}} \quad (4)$$

其中  $\sigma$  表示窗口的宽度。不同的数据集对应于不同的  $\sigma$  值, 而且  $\sigma$  的大小直接影响着概率密度估计的效果。对于该参数, 本文采用基于相似度的交叉验证法(cross validation)来确定  $\sigma$  的值<sup>[17]</sup>。

根据(4)式, 对于数据集  $S_1$  和  $S_2$ , 可以分别得出概率密度函数  $f_1(\mathbf{z}, \sigma)$  和  $f_2(\mathbf{z}, \sigma)$ :

$$f_1(\mathbf{z}, \sigma) = \frac{1}{N} \sum_{i=1}^N (\sqrt{2\pi}\sigma)^{-d} e^{-\frac{\|\mathbf{z}-\mathbf{z}'_i\|^2}{2\sigma^2}} \quad (5)$$

$$f_2(\mathbf{z}, \sigma) = \frac{1}{N} \sum_{i=1}^N (\sqrt{2\pi}\sigma)^{-d} e^{-\frac{\|\mathbf{z}-\mathbf{z}''_i\|^2}{2\sigma^2}} \quad (6)$$

令  $G(\mathbf{z}-\mathbf{z}', \sigma^2) = e^{-\frac{\|\mathbf{z}-\mathbf{z}'\|^2}{2\sigma^2}}$ , 则(5)式和(6)式可化为:

$$f_1(\mathbf{z}, \sigma) = \frac{(\sqrt{2\pi}\sigma)^{-d}}{N} \sum_{i=1}^N G(\mathbf{z}-\mathbf{z}'_i, \sigma^2) \quad (7)$$

$$f_2(\mathbf{z}, \sigma) = \frac{(\sqrt{2\pi}\sigma)^{-d}}{N} \sum_{i=1}^N G(\mathbf{z}-\mathbf{z}''_i, \sigma^2) \quad (8)$$

利用上节中提到的相对熵的性质可知, 相对熵越大, 两个概率密度分布的差别就越小, 当相对熵达到最大值时, 两个概率密度分布完全相同, 即  $f_2(\mathbf{z}, \sigma)$  和  $f_1(\mathbf{z}, \sigma)$  同分布。换言之, 网络的预测输出  $y_i$  接近样本输出  $y_{di}$ 。为此, 定义如下新的目标函数:

$$\begin{aligned} E_2 &= -W(f_1, f_2) \\ &= -[V(f_1, f_2) + V(f_2, f_1)] \\ &= -\int f_1(\mathbf{z}) \log[f_1(\mathbf{z})/f_2(\mathbf{z})] d\mathbf{z} + \int f_2(\mathbf{z}) \log[f_2(\mathbf{z})/f_1(\mathbf{z})] d\mathbf{z} \\ &= -\int \{f_1(\mathbf{z})[\log f_1(\mathbf{z}) - \log f_2(\mathbf{z})] + f_2(\mathbf{z})[\log f_2(\mathbf{z}) - \log f_1(\mathbf{z})]\} d\mathbf{z} \end{aligned} \quad (9)$$

根据 Taylor 级数展开,  $\log f(\mathbf{z})$  可以近似为:

$$\log f(\mathbf{z}) = \log[1 + (f(\mathbf{z}) - 1)] \approx f(\mathbf{z}) - 1 \quad (10)$$

将(10)代入(9)化简得

$$E_2 = \int [f_1^2(\mathbf{z}) + f_2^2(\mathbf{z}) - 2f_1(\mathbf{z})f_2(\mathbf{z})] d\mathbf{z} \quad (11)$$

在参考文献[18]中, Erhan 和 Jose 详细推导了如下结论:

$$\int f_1^2(\mathbf{z}) d\mathbf{z} = \frac{(\sqrt{2\pi}\sqrt{2}\sigma)^{-d}}{N^2} \sum_{i_1=1}^N \sum_{i_2=1}^N G(\mathbf{z}'_{i_1} - \mathbf{z}'_{i_2}, 2\sigma^2) \quad (12)$$

$$\int f_2^2(\mathbf{z}) d\mathbf{z} = \frac{(\sqrt{2\pi}\sqrt{2}\sigma)^{-d}}{N^2} \sum_{i_1=1}^N \sum_{i_2=1}^N G(\mathbf{z}''_{i_1} - \mathbf{z}''_{i_2}, 2\sigma^2) \quad (13)$$

$$\int f_1(\mathbf{z})f_2(\mathbf{z}) d\mathbf{z} = \frac{(\sqrt{2\pi}\sqrt{2}\sigma)^{-d}}{N^2} \sum_{i_1=1}^N \sum_{i_2=1}^N G(\mathbf{z}'_{i_1} - \mathbf{z}''_{i_2}, 2\sigma^2) \quad (14)$$

将(12), (13), (14)代入(11)化简可得:

$$\begin{aligned} E_2 &= \frac{(\sqrt{2\pi}\sqrt{2}\sigma)^{-d}}{N^2} \left[ \sum_{i_1=1}^N \sum_{i_2=1}^N G(\mathbf{z}'_{i_1} - \mathbf{z}'_{i_2}, 2\sigma^2) \right. \\ &\quad \left. + \sum_{i_1=1}^N \sum_{i_2=1}^N G(\mathbf{z}''_{i_1} - \mathbf{z}''_{i_2}, 2\sigma^2) - 2 \sum_{i_1=1}^N \sum_{i_2=1}^N G(\mathbf{z}'_{i_1} - \mathbf{z}''_{i_2}, 2\sigma^2) \right] \end{aligned} \quad (15)$$

如上所述, 我们为 RBF 建模构造了一个基于概率密度函数和熵准则的新的准则函数, 该准则函数能暗含样本集的潜在结构, 从而有效地克服了传统的 MSE 准则函数只能从局部进行学习的缺陷。利用梯度优化方法和提出的新的基于熵准则的目标函数, 容易推得 RBF 网络参数学习的新规则(详细推导见附录)。

### 3 实验结果与分析

#### 3.1 性能指标

对神经网络而言, 网络训练后对测试样本集的响应如何, 直接决定了网络的性能。为了方便比较, 采用如下常用的性能指标来评价模型的泛化能力和鲁棒性<sup>[19,20]</sup>:

$$J = \sqrt{\frac{\sum_{l=1}^N (y_l - \bar{y})^2}{\sum_{l=1}^N (y_l^d - \bar{y})^2}} \quad (16)$$

其中,  $\bar{y} = \frac{1}{N} \sum_{l=1}^N y_l^d$ ,  $N$  为样本总数,  $y_l^d$  为第  $l$  个测试样本值的采样输出,  $y_l$  为测试样本的网络模型输出。J 值越小, 意味着模型的预测精度越高, 泛化能力越强。

#### 3.2 结果与分析

如表 2 所示, 试验数据中对应的 5 个因素水平值为 RBF 网络的输入, GSH 的浓度作为网络的输出。因为实际的发酵过程非常复杂, 试验数据往往

不可避免的带有一定量的噪音，所以在进行发酵建模过程中，模型的鲁棒性是一个非常重要的指标。因此，我们对试验数据累计加入了 10 次不同程度的噪音，每次噪音的增量为 0.02，分别采用式(1)和(15)所示的两种误差准则函数对 RBF 网络进行训练，考察两种方法的预测精度及鲁棒性。

根据数据集的特点，网络有 5 个输入，1 个输出。通过交叉验证法<sup>[21]</sup>分析得到，当隐层神经元的个数取 9~12 个时，能够得到比较满意的结果。为了节省篇幅，文中仅给出了神经元个数为 10 时的实验结果和实验分析。另外，我们设定相同的收敛标准，即在梯度下降算法过程中，当梯度小于  $10^{-4}$  时，算法收敛，停止迭代。下面采用 Jackknife 检验方法对模型的泛化性能进行分析，即对表 2 中的 16 组试验数据，每次从中取出 15 组作为训练集，剩余的一组数据用于预测，依次循环 16 次，这样每一组数据都得到了一个预测值，根据式(16)定义的衡量模型的泛化能力及鲁棒性的指标，对模型的性能进行评价，实验结果如下表 3 所示。由于篇幅有限，这里我们仅给出了部分实验的网络训练后的参数，见表 4 和表 5 所示的噪音量为 0.00 以及 0.10 时，训练得到的网络参数值。

从上表 3 中对 GSH 浓度的预测结果可以看出：当试验数据没有加入噪音时(如表 3 中的第一行数据所示)，基于熵准则的模型的  $J$  值为 0.6782，而传统的基于 MSE 准则的模型的  $J$  值为 0.8299，这说明基

于熵准则的模型的预测精度高，预测能力强。随着数据中的噪音从 0.02 增加到 0.20 的过程中，基于熵准则的 RBF 网络模型对应的  $J$  值从 0.7021 增加到 0.7444，而传统的基于 MSE 准则的 RBF 模型的  $J$  值则从 0.8448 逐步递增到 1.2065。即本文提出的基于熵准则的模型对应的  $J$  值始终小于传统的基于 MSE 准则的模型的  $J$  值，而且变化相对比较平缓，这就说明当试验数据存在噪音的情况下，新模型的预测能力仍然高于传统模型，因此新模型具有较好的泛化能力和鲁棒性。下面分别以 DCW 的浓度和胞内 GSH 的浓度作为网络的输出，其网络参数的设置及实验过程均同上，详细实验结果如表 6、表 7 所示。

表 3 正交试验中 GSH 浓度预测性能比较  
Table 3 Comparison of GSH concentration prediction of the orthogonal experiments

No.	Noise	Performance index $J$	
		RBF networks based MSE	RBF networks based Entropy
1	0.00	0.8299	0.6782
2	0.02	0.8448	0.7021
3	0.04	0.8599	0.7065
4	0.06	0.8708	0.7136
5	0.08	0.8987	0.7184
6	0.10	0.9366	0.7137
7	0.12	0.9887	0.7127
8	0.14	1.0669	0.7190
9	0.16	1.1154	0.7184
10	0.18	1.1964	0.7251
11	0.20	1.2065	0.7444

表 4 噪音量为 0.00 时训练得到的网络参数  
Table 4 The network parameters after training with the noise at 0.00

Hidden nodes	RBF networks based MSE criterion		RBF networks based entropy criterion	
	Weight	Center	Weight	Center
1	0.546480	(0.898950, 0.162280, 0.474550, 1.227400, 0.097299)	0.603600	(0.023000, 0.663000, 0.680800, 0.680100, 0.665600)
2	-0.404150	(0.673400, 0.235730, 1.160100, 0.755980, -0.176087)	0.982400	(0.000600, -0.013300, 0.015200, -0.002100, 0.018900)
3	1.063400	(-0.102870, 0.089856, 0.230230, -0.662260, 0.083226)	0.479300	(0.357100, -0.005700, 0.314100, 0.662900, 0.982500)
4	1.118900	(-0.096623, 0.765440, 1.008200, 0.574690, 0.792860)	0.525300	(-0.005100, 1.008000, 1.001900, 0.988600, 1.017000)
5	-0.068189	(0.102900, -0.115440, 1.142500, 0.330230, 0.672910)	0.443300	(-0.311600, 0.667900, 0.996700, -0.010700, 0.327900)
6	0.233170	(0.408540, 0.995490, 0.763360, 0.251470, -0.021160)	0.388000	(0.09700, 0.720800, 0.338700, 0.994500, 0.011300)
7	-0.169220	(0.734240, 1.151900, 0.217970, -0.045359, 0.723530)	-0.148500	(0.662600, 0.665500, -0.007800, 0.344400, 0.988400)
8	0.554810	(0.194670, -0.239240, 0.323700, 0.685560, 1.184200)	-0.330600	(0.670700, 0.331900, 1.014900, 0.681200, 0.002300)
9	0.057640	(0.306350, 0.347890, -0.178490, 1.171800, 0.637480)	0.224700	(0.625400, -0.016100, 0.657600, 0.993200, 0.368400)
10	0.035118	(0.108500, 0.316950, 0.679540, -0.124260, 1.131700)	0.131200	(0.336000, 0.995300, 0.679400, 0.327300, -0.022700)

表 5 噪音量为 0.10 时训练得到的网络参数  
Table 5 The network parameters after training with the noise at 0.10

Hidden nodes	RBF networks based MSE criterion		RBF networks based entropy criterion	
	Weight	Center	Weight	Center
1	-0.154180	(0.666150, 0.710200, -0.139500, 0.223140, 1.188500)	0.594300	(-0.021900, 0.655100, 0.634400, 0.644400, 0.651300)
2	0.535370	(0.340390, -0.263000, 0.441820, 0.734570, 1.091000)	0.869200	(0.048600, 0.073800, -0.017900, -0.061300, -0.041800)
3	0.976200	(-0.213330, 0.827210, 0.839730, 0.769730, 0.814290)	0.468400	(0.337900, 0.018300, 0.359200, 0.639700, 1.026600)
4	0.206410	(0.387460, 1.171700, 0.661150, 0.382630, -0.162940)	0.441600	(0.295800, 0.653300, 0.974600, -0.036100, 0.349700)
5	0.078432	(1.168600, 0.264760, 0.700420, -0.094746, 1.151500)	0.525500	(0.003400, 0.978800, 0.986800, 1.033000, 0.959200)
6	0.339560	(0.155380, 0.657690, 1.194900, -0.205470, 0.427930)	0.333600	(1.028400, 0.673700, 0.267800, 0.972500, 0.002500)
7	0.960220	(-0.135330, 0.023643, 0.043977, 0.035000, 0.043829)	-0.145800	(0.643700, 0.678800, 0.025500, 0.329600, 1.040600)
8	0.037310	(0.778740, 0.333190, 0.980850, 0.871030, -0.185360)	0.181700	(0.054100, 0.349100, 0.345000, 0.343200, 0.266500)
9	-0.166320	(1.113500, -0.154050, 1.148500, 0.319980, 0.645980)	-0.255200	(0.708000, 0.363000, 1.024500, 0.666500, 0.012900)
10	0.146070	(0.326450, 0.389280, -0.095472, 1.172500, 0.549120)	0.226300	(0.655200, -0.019900, 0.631500, 0.931700, 0.309200)

表 6 正交试验中 DCW 浓度预测性能比较  
Table 6 Comparison of GSH concentration prediction of the orthogonal experiments

No.	Noise	Performance index J	
		RBF networks based MSE	RBF networks based Entropy
1	0.00	0.8299	0.6978
2	0.02	0.8471	0.7087
3	0.04	0.8761	0.7268
4	0.06	0.8989	0.7362
5	0.08	0.9198	0.7381
6	0.10	0.9443	0.7380
7	0.12	0.9529	0.7368
8	0.14	1.0127	0.7497
9	0.16	1.0606	0.7729
10	0.18	1.0818	0.7765
11	0.20	1.1020	0.7841

表 7 正交试验中胞内 GSH 浓度预测性能比较  
Table 7 Comparison of Intracellular GSH content prediction of the orthogonal experiments

No.	Noise	Performance index J	
		RBF networks based MSE	RBF networks based Entropy
1	0.00	0.7267	0.7149
2	0.02	0.7499	0.7178
3	0.04	0.7885	0.7131
4	0.06	0.8126	0.7240
5	0.08	0.8225	0.7218
6	0.10	0.8612	0.7230
7	0.12	0.8843	0.7433
8	0.14	0.9287	0.7573
9	0.16	0.9655	0.7746
10	0.18	1.0544	0.7773
11	0.20	1.1200	0.7820

显然,从表 6 和表 7 可以得出相同的结论,故本文中所提出的基于熵准则的 RBF 建模方法,其预测精度明显的要高于传统的基于 MSE 准则的 RBF 建模方法,能更准确地预测出谷胱甘肽的发酵过程中各成分的浓度,显示出了更好的泛化能力,即具有更好的鲁棒性。

为了更直观地显示出本文中提出的新方法的良好鲁棒性,我们以噪音为横坐标轴,预测误差指数为纵坐标轴,这两种方法的效果对比如图 1~3 所示。

从图 1、图 2 和图 3 可以看出:在对 GSH 浓度、DCW 浓度以及胞内 GSH 浓度的预测中,基于熵准则方法对应的误差指数曲线始终位于传统的基于 MSE 准则方法的曲线的下部,而且曲线的变化相对比较平稳,这也有效地证实了本文中提出的新方法具有更好的预测精度、泛化能力和鲁棒性。

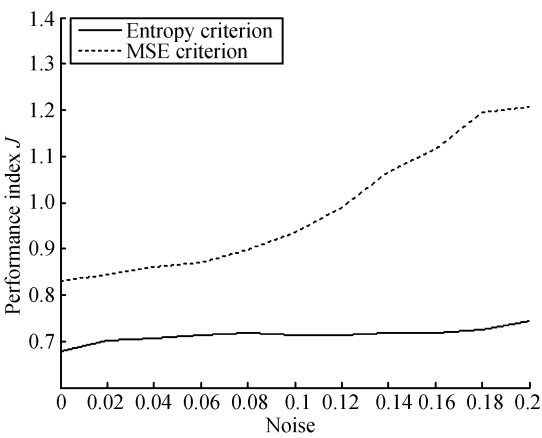


图 1 GSH 浓度预测比较  
Fig. 1 Comparison of GSH concentration prediction

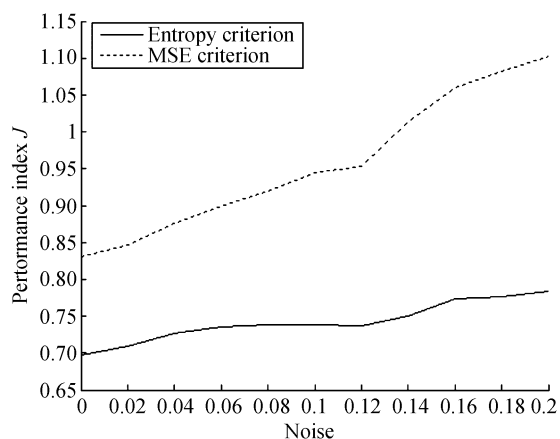


图 2 DCW 浓度预测比较

Fig. 2 Comparison of DCW concentration prediction

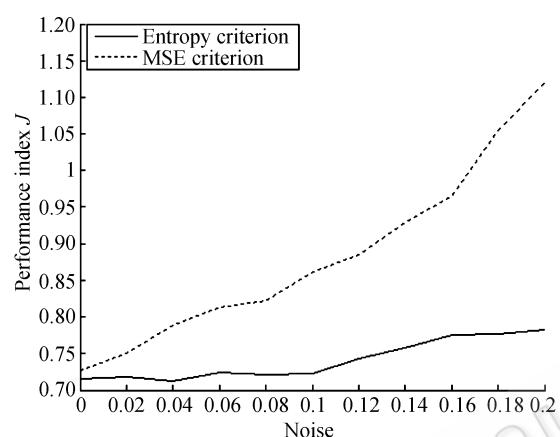


图 3 胞内 GSH 浓度预测比较

Fig. 3 Comparison of Intracellular GSH content prediction

## 4 讨论

在谷胱甘肽的发酵过程中, 由于试验仪器以及条件的限制, 所得到的试验数据往往不可避免地带有噪音, 传统的基于 MSE 准则的神经网络模型是一个局部逼近过程, 因此学习所得的网络的泛化能力和鲁棒性较差。本文从一个新的角度提出了基于熵准则的 RBF 神经网络模型, 它能够从全局出发, 在参数学习阶段考虑了所有样本的整体分布结构。运用此模型对谷胱甘肽的试验数据进行建模, 结果证明新方法具有较高的预测精度、泛化能力和良好的鲁棒性。这对于实际的谷胱甘肽的发酵生产中, 在试验数据由于客观条件的限制往往含有一定噪音的情况下的建模问题, 提供了一种新方法, 具有一定的潜在应用价值。

## 附录

利用提出的熵准则函数(如式(15)所示)和梯度学习算法, 我们很容易得到如下的 RBF 神经网络的参数学习规则:

$$p_j^i(t+1) = p_j^i(t) - r_1 \frac{\partial E_2}{\partial p_j^i} \quad (17)$$

$$a_j^i(t+1) = a_j^i(t) - r_2 \frac{\partial E_2}{\partial a_j^i} \quad (18)$$

$$b_j^i(t+1) = b_j^i(t) - r_2 \frac{\partial E_2}{\partial b_j^i} \quad (19)$$

其中训练次数  $t = 0, 1, 2, \dots$ ;  $r_1, r_2$  是学习率;  $i=1, 2, \dots, M$ ;  $j = 1, 2, \dots, d$ 。这里,

$$\frac{\partial E_2}{\partial p_j^i} = -\frac{(\sqrt{2\pi}\sqrt{2\sigma})^{-d}}{2\sigma^2 N^2} \left[ \sum_{t_1=1}^N \sum_{t_2=1}^N G(\mathbf{z}_{t_1}'' - \mathbf{z}_{t_2}'', 2\sigma^2)(y_{t_1} - y_{t_2}) \cdot \left( \frac{\partial y_{t_1}}{\partial p_j^i} - \frac{\partial y_{t_2}}{\partial p_j^i} \right) + 2 \sum_{t_1=1}^N \sum_{t_2=1}^N G(\mathbf{z}_{t_1}' - \mathbf{z}_{t_2}'', 2\sigma^2)(y_{dt_1} - y_{t_2}) \frac{\partial y_{t_2}}{\partial p_j^i} \right] \quad (20)$$

$$\frac{\partial y_{t_1}}{\partial p_j^i} = \frac{w^i}{\sum_{t_3=1}^M w^{t_3}} x_{t_1j} \quad (21)$$

$$\frac{\partial y_{t_2}}{\partial p_j^i} = \frac{w^i}{\sum_{t_3=1}^M w^{t_3}} x_{t_2j} \quad (22)$$

$$\frac{\partial E_2}{\partial a_j^i} = -\frac{(\sqrt{2\pi}\sqrt{2\sigma})^{-d}}{2\sigma^2 N^2} \left[ \sum_{t_1=1}^N \sum_{t_2=1}^N G(\mathbf{z}_{t_1}'' - \mathbf{z}_{t_2}'', 2\sigma^2)(y_{t_1} - y_{t_2}) \cdot \left( \frac{\partial y_{t_1}}{\partial a_j^i} - \frac{\partial y_{t_2}}{\partial a_j^i} \right) + 2 \sum_{t_1=1}^N \sum_{t_2=1}^N G(\mathbf{z}_{t_1}' - \mathbf{z}_{t_2}'', 2\sigma^2)(y_{dt_1} - y_{t_2}) \frac{\partial y_{t_2}}{\partial a_j^i} \right] \quad (23)$$

$$\frac{\partial y_{t_1}}{\partial a_j^i} = \frac{y_{t_1}^i \sum_{t_3=1}^M w^{t_3} - \sum_{t_3=1}^M (w^{t_3} y_{t_1}^{t_3})}{(\sum_{t_3=1}^M w^{t_3})^2} \cdot \frac{2 \cdot (x_{t_1j} - a_j^i)}{b_j^i} \cdot w^i \quad (24)$$

$$\frac{\partial y_{t_2}}{\partial a_j^i} = \frac{y_{t_2}^i \sum_{t_3=1}^M w^{t_3} - \sum_{t_3=1}^M (w^{t_3} y_{t_2}^{t_3})}{(\sum_{t_3=1}^M w^{t_3})^2} \cdot \frac{2 \cdot (x_{t_2j} - a_j^i)}{b_j^i} \cdot w^i \quad (25)$$

$$\frac{\partial E_2}{\partial b_j^i} = -\frac{(\sqrt{2\pi}\sqrt{2\sigma})^{-d}}{2\sigma^2 N^2} \left[ \sum_{t_1=1}^N \sum_{t_2=1}^N G(\mathbf{z}_{t_1}'' - \mathbf{z}_{t_2}'', 2\sigma^2)(y_{t_1} - y_{t_2}) \cdot \left( \frac{\partial y_{t_1}}{\partial b_j^i} - \frac{\partial y_{t_2}}{\partial b_j^i} \right) + 2 \sum_{t_1=1}^N \sum_{t_2=1}^N G(\mathbf{z}_{t_1}' - \mathbf{z}_{t_2}'', 2\sigma^2)(y_{dt_1} - y_{t_2}) \frac{\partial y_{t_2}}{\partial b_j^i} \right] \quad (26)$$

$$\frac{\partial y_{t_1}}{\partial b_j^i} = \frac{y_{t_1}^i \sum_{t_3=1}^M w^{t_3} - \sum_{t_3=1}^M (w^{t_3} y_{t_1}^{t_3})}{(\sum_{t_3=1}^M w^{t_3})^2} \cdot \frac{(x_{t_1j} - a_j^i)^2}{(b_j^i)^2} \cdot w^i \quad (27)$$

$$\frac{\partial y_{t_2}}{\partial b_j^i} = \frac{y_{t_2}^i \sum_{t_3=1}^M w^{t_3} - \sum_{t_3=1}^M (w^{t_3} y_{t_2}^{t_3})}{(\sum_{t_3=1}^M w^{t_3})^2} \cdot \frac{(x_{t_2j} - a_j^i)^2}{(b_j^i)^2} \cdot w^i \quad (28)$$

## REFERENCES

- [1] Meister A, Anderson ME. Glutathione. *Ann Rev Biochem*, 1983, **52**: 711–760.
- [2] Izawa S, Inoue Y, Kimura A. Oxidative stress response in yeast: effect of glutathione on adaptation to hydrogen peroxide stress in *Saccharomyces cerevisiae*. *FEBS Lett*, 1995, **368**: 73–76.
- [3] Meister A. Antioxidant functions of glutathione. *Life Chem Rep*, 1994b, **12**(1): 23–27.
- [4] Jahoor F, Jackson A, Gazzard B, Philips G, Sharpstone D, Frazer ME, Heird W. Erythrocyte glutathione deficiency in symptom-free HIV infection is associated with decreased synthesis rate. *Am J Physiol*, 1999, **276**(1): E205–E211.
- [5] Linko P, Zhu YH. Neural network modeling for real time variable estimation and prediction in the control of glucoamylase fermentation. *Process Biochem*, 1992, **27**: 275–283.
- [6] Kennedy MJ, Prapulla SG, Thakur MS. A comparison of neural networks to factorial design. *Biotechnol Tech*, 1992, **6**: 293–299.
- [7] Desai KM, Akolkar SK, Badhe YP, Tambe SS, Lele SS. Optimization of fermentation media for exopolysaccharide production from *Lactobacillus plantarum* using artificial intelligence based techniques. *Process Biochemistry*, 2006, **41**: 1842–1848.
- [8] Zhang GY, Fang BS. A model for amino acid composition and optimum pH in G/11 xylanase based on neural networks. *Chinese Journal of Biotechnology*, 2005, **21**(4): 658–661.  
张光亚, 方柏山. 木聚糖酶氨基酸组成与其最适 pH 的神经网络模型. *生物工程学报*, 2005, **21**(4): 658–661.
- [9] Lin Y, Cai FY, Zhang GY. A prediction model for the activity of insecticidal crystal proteins from *Bacillus thuringiensis* based on support vector machine. *Chinese Journal of Biotechnology*, 2007, **23**(1): 127–132.  
林毅, 蔡福营, 张光亚. 苏云金杆菌杀虫晶体蛋白活性预测的支持向量机模型. *生物工程学报*, 2007, **23**(1): 127–132.
- [10] Lin Y, Cai FY, Yuan YX, Zhang GY. A uniform design based PCA-SVM model for predicting optimum pH in chitinase. *Chinese Journal of Biotechnology*, 2007, **23**(3): 514–519.  
林毅, 蔡福营, 袁宇熹, 张光亚. 基于均匀设计的主成分分析-支持向量机模型及其在几丁质酶最适 pH 建模中的应用. *生物工程学报*, 2007, **23**(3): 514–519.
- [11] Wang J, Chen L, Yang HJ, Tan QQ, Liu SY, Zhang KX. Fermentation process of L-tryptophan modeling based on BP neural networks. *Journal of Tianjin University of Light Industry*, 2003, **18**(3): 1–4.  
王健, 陈宁, 杨海军, 谭青乔, 刘淑云, 张克旭. 基于 BP 神经网络的 L-色氨酸发酵过程建模. *天津轻工业学院学报*, 2003, **18**(3): 1–4.
- [12] Bian ZQ, Zhang XG, *et al.* Pattern Recognition. Beijing: Press of Tsing-hua University, 1999.  
边肇祺, 张学工, 等. 模式识别. 北京: 清华大学出版社, 1999.
- [13] Teaching and Research Group of the School of Biochemistry in Peking University. Experimental Instructions of Biochemistry. Beijing: Higher Education Press, 1984, 22–24.  
北京大学生物系生物化学教研室编. 生物化学实验指导. 北京: 高等教育出版社, 1984, 22–24.
- [14] Tietze F. Enzymic method for quantitative determination of nanogram amounts of total and oxidized glutathione: application to mammalian blood and other tissues. *Anal Biochem*, 1969, **27**: 502–522.
- [15] Wong CC, Lai HR. Generating fuzzy control rules by a clustering algorithm based on a grey relational measure. *Proc of IEEE Int Conf Fuzzy Systems*. Seoul Korea: IEEE Press, 1999, 470–473.
- [16] Zhao ZY, Xu YM. The fundamentals and applications of fuzzy theories and neural networks. Beijing: Press of Tsing-hua University, 1996.  
赵振宇, 徐用懋. 模糊理论和神经网络的基础与应用. 北京: 清华大学出版社, 1996.
- [17] Girolami M, He Chao. Probability density estimation from optimally condensed data samples. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2003, **25**(10): 1253–1264.
- [18] Erhan G, Jose CP. Information Theoretic Clustering. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2002, **24**(2): 158–171.
- [19] Jang SR, Sun CT, Mizutani E. Neuro-Fuzzy and Soft Computing. Upper Saddle River, NJ: Prentice-Hall, 1997.
- [20] Chung FL, Wang ST, Deng ZH, Hu DW. CATSMLP: Towards a robust and interpretable multilayer perceptron with sigmoid activation functions. *IEEE Transactions on Systems, Man and Cybernetics-Part B*, 2006, **36**(6): 1319–1331.
- [21] Kearns M, Ron D. Algorithmic stability and sanity-check bounds for leave-one-out cross-validation, *Tenth Annual Conference on Computational Learning Theory*, 1997, 152–162.
- [22] Wei GY, Li Y, Du GC, Chen J. Fermentation conditions of glutathione by *Candida utilis*. *Appl Environ Biol*, 2003, **9**(6): 642–646.  
卫功元, 李寅, 堵国成, 陈坚. *Candida utilis* 生物合成谷胱甘肽的营养及环境条件. *应用与环境生物学报*, 2003, **9**(6): 642–646.