J. Biomedical Science and Engineering, 2010, 3, 430-438 JBiSE
doi:10.4236/jbise.2010.34059 Published Online April 2010 (http://www.SciRP.org/journal/jbise/).

Fermentation process modeling of exopolysaccharide using neural networks and fuzzy systems with entropy criterion

Zuo-Ping Tan[1,2,3], Shi-Tong Wang[1], Zhao-Hong Deng[1], Guo-Cheng Du[2]

[1]School of Information Engineering, Jiangnan University, Wuxi, china;
[2]Key Laboratory of Industrial Biotechnology, Ministry of Education, Jiangnan University, Wuxi, China;
[3]Jiangsu Province Key Laboratory of Information Technology at Suzhou University, Suzhou, China.
Email: wxwangst@yahoo.com.cn

Received 17 December 2009; revised 28 December 2009; accepted 12 January 2010.

.

ABSTRACT

The prediction accuracy and generalization of fermentation process modeling on exopolysaccharide (EPS) production from Lactobacillus are often deteriorated by noise existing in the corresponding experimental data. In order to circumvent this problem, a novel entropy-based criterion is proposed as the objective function of several commonly used modeling methods, *i.e.* Multi-Layer Perceptron (MLP) network, Radial Basis Function (RBF) neural network, Takagi-Sugeno-Kang (TSK) fuzzy system, for fermentation process model in this study. Quite different from the traditional Mean Square Error (MSE) based criterion, the novel entropy-based criterion can be used to train the parameters of the adopted modeling methods from the whole distribution structure of the training data set, which results in the fact that the adopted modeling methods can have global approximation capability. Compared with the MSE-criterion, the advantage of this novel criterion exists in that the parameter learning can effectively avoid the over-fitting phenomenon, therefore the proposed criterion based modeling methods have much better generalization ability and robustness. Our experimental results confirm the above virtues of the proposed entropy-criterion based modeling methods.

Keywords: Relative Entropy; MSE-Criterion Based Modeling; Robustness; Parzen Window; TSK Fuzzy System

1. INTRODUCTION

Polysaccharides are produced by plants, algae and bacteria, which are used in pharmaceutical, chemical, pesticide and oil exploitation. Some microorganisms such as the lactic acid producers are known to synthesize exopolysaccharides (EPS), which can be used commercially as food additives and have health stimulating properties such as immunity stimulation, anti-ulcer activity and cholesterol reduction. However, as we may know well, EPS's fermentation mechanism is very complex because it refers to the growth and reproduction of microorganisms [1]. In view of control, fermentation process contains high non-linearity, high time-varying and uncertainty. Meanwhile the lack of biosensor and the interaction of coupled parameters also bring much difficulty for the fermentation process modeling [2]. In the last decade, artificial neural networks (ANNs) have been proved to be able to model nonlinear systems and successfully applied in various chemical and biological models [3]. Especially they have emerged as an attractive tool for predicting and approximating the parameters in fermentation process [4], and demonstrated their powers in the factorial design [5]. More examples include one ANNs-based model for amino acid composition and optimum pH in G/11 xylanase [6], and another ANNs-based model for optimization of fermentation media for exopolysaccharide production from Lactobacillus plantarum [7]. In recent years, fuzzy systems and/or fuzzy neural networks researchers have paid particular attention on industrial fermentation process modeling [8]. For instance, fuzzy neural network has been used for dissolved Oxygen predictive control of fermentation process [9], and Takagi-Sugeno-Kang (TSK) fuzzy system has been used for biochemical variable estimation of fermentation process [10]. In addition, an application of fuzzy control in citric acid fermentation process has been adopted to maximize the biomass quantities [11]. However, when MSE-criterion based objective function is used for model parameter learning, the above methods have the so-called over-fitting drawback, that is to say, MSE-criterion based modeling methods may over-fit each training sample such that the whole distribution of the training set is erroneously estimated and the generalization ability can not be assured.

In this study, in order to overcome the weaknesses mentioned above, the new criterion is proposed as the objective function for fermentation process modeling. This new criterion, called the entropy criterion, is based on the probability density estimation for the whole training set and relative entropy [12]. And then the proposed criterion is used in the classical Multi- Layer Perceptron (MLP) network modeling, Radial Basis Function (RBF) neural network modeling and Takagi-Sugeno-Kang (TSK) fuzzy system modeling, for the EPS fermentation process modeling.

2. MATERIALS AND METHODS

The data we used in this study was derived from the reference [7]. This project was conducted in 2004-2006 by Mumbai University of Food Engineering, in Mumbai, India.

2.1. Bacterial Strain

Lactobacilli strain is isolated from the Indian fermented food ragi. This isolation is characterized as Lactobacillus plantarum using biochemical tests.

2.2. Medium

The medium contain lactose, casine hydrolysate, triammonium citrate, beef extract and proteose peptone, along with sodium acetate: 1 g/l, Mg-sulfate: 1 g/l, manganese sulfate: 0.5 g/l and calcium chloride: 0.25 g/l. The medium are autoclaved at 110℃ for 10 min; lactose is autoclaved separately.

2.3. Fermentation Conditions

The batch fermentation is carried in a 250 ml shake flask for 24 h at 150 rpm and 35℃. The pH of the fermentation medium is adjusted to 6.5 ± 0.3 with the addition of 1N NaOH/1N HCL. Flasks at the end of fermentation are analyzed for EPS production.

2.4. Analysis

The cells are separated by centrifugation(10,000 rpm, 10℃, 15 min) and the crude EPS is precipitated from the broth at 4 by the addition of two volumes of cold ethanol (95%). The resulting precipitate is collected by centrifugation and re-dissolved in water. The crude EPS solution is dialyzed at 4℃ to estimate the yield.

2.5. MSE-Criterion Based Fermentation Process Modeling

In most of current modeling methods, the MSE-criterion based objective function is often used for training model parameters. The MSE-criterion can be formulated as

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

where y_i, y_{di} are the predicted and desired output for ith sample , respectively.

From **Eq.1**, we can see that the MSE-criterion based model parameter learning is just a local approximation process and does not consider the whole distribution of the training set [13,14], thus the generalization and robustness of the model will not be ensured and the over-fitting often occurs, especially when there are noises in the training data.

3. ENTROPY-CRITERION BASED FERMENTATION PROCESS MODELING

3.1. Relative Entropy and Jeffreys-Divergence Entropy

Entropy is a measurement of uncertainty in information theory, which is a function of the probability density distribution. The concept of relative entropy can be introduced to measure the difference between certain probability density distribution $f_1(x_i)$ and a given probability density distribution $f_2(x_i)$, which may be written as follows [12,15],

$$V(f_1, f_2) = -\sum f_1(x_i)\log[f_1(x_i)/f_2(x_i)] \le 0 \tag{2}$$

where the smaller value of relative entropy is, the larger difference between the two density distributions is. Meanwhile, when certain probability density distribution is equal to the given distribution, the relative entropy will reach its maximum (equal to zero). It is well known that relative entropy is additive and non-symmetrical. To obtain a symmetrical measure, Jeffreys-divergence entropy (J-divergence entropy) can be used. It is also called symmetrical relative entropy which can measure the difference between two densities $f_1(x_i)$ and $f_2(x_i)$.

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

According to the above J-divergence entropy, a novel objective function based on entropy-criterion will be illustrated in the next subsection.

3.2. Relative Entropy Based Objective Function

For a given training sample set $\{(\mathbf{x}_i, y_{di}) | \mathbf{x}_i \in R^d,\ y_{di} \in R,\ i = 1,2,...,N\}$, we can reconstruct two new sets, *i.e.*, one contains the sample inputs and the sample outputs, $S_1 = \{\mathbf{z}_i' | \mathbf{z}_i' = (\mathbf{x}_i, y_{di})\}$,

$\mathbf{z}_i^{'} \in R^{d^{'}}$ ($d^{'} = d+1$), and the other contains the sample inputs and the model predicted outputs, $S_2 = \{\mathbf{z}_i^{''} | \mathbf{z}_i^{''} = (\mathbf{x}_i, y_i)\}$, $\mathbf{z}_i^{''} \in R^{d^{'}}$.

For the above sample set $\{\mathbf{x}_i | \mathbf{x}_i \in R^d, \ i = 1,2,...,N\}$, its probability density can be estimated with the following parzen window density estimator,

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

where σ represents the window width. For a given data, it is constant and can be used to effectively estimate the corresponding density distribution. Here with maximum likelihood estimation (MLE), it is determined through cross-validation (CV) method, and the value resulting in the max magnitude is chosen [16].

For the above two data sets S_1 and S_2, their probability density distribution functions, $f_1(\mathbf{z},\sigma)$ and $f_2(\mathbf{z},\sigma)$, can be formulated as

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

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

Suppose $G(\mathbf{z}-\mathbf{z}^{'},\sigma^2) = e^{-\frac{\|\mathbf{z}-\mathbf{z}_i^{'}\|}{2\sigma^2}}$, then we get,

$$f_1(\mathbf{z},\sigma) = \frac{(\sqrt{2\pi}\sigma)^{-d^{'}}}{N}\sum_{t_1=1}^{N}G(\mathbf{z}-\mathbf{z}^{'},\sigma^2) \tag{7}$$

$$f_2(\mathbf{z},\sigma) = \frac{(\sqrt{2\pi}\sigma)^{-d^{'}}}{N}\sum_{t_2=1}^{N}G(\mathbf{z}-\mathbf{z}^{''},\sigma^2) \tag{8}$$

By using the properties of relative entropy, the bigger the value of relative entropy is, the smaller the difference between two probability densities is, as aforementioned. When the relative entropy reaches its maximal value, the two density functions will absolutely be the same, *i.e.*, $f_2(\mathbf{z},\sigma)$ is equal to $f_1(\mathbf{z},\sigma)$. In other words, in this case the predicted output y_i of the model approximates the sample output y_{di} in the training set well. Consequently the novel objective function may be defined as

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

From **Eq.4-Eq.6**, we can see that f(z) is obtained by Parzen window estimator , thus its value ranges from 0 to 1. According to the properties of Taylor's expansion, when $f(\mathbf{z})$ is small, we can just keep the linear parts of $\log f(\mathbf{z})$, that is to say, $\log f(\mathbf{z})$ can be simplified as follows,

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

Therefore, submitting **Eq.10** into **Eq.9**, we get,

Please note, Erhan and Jose [17] have strictly inferred the following formulas,

$$\int f_1^2(\mathbf{z})\,d\mathbf{z} = \frac{\left(\sqrt{2\pi}\sqrt{2}\sigma\right)^{-d'}}{N^2}\sum_{t_1=1}^{N}\sum_{t_2=1}^{N}G\left(\mathbf{z}'_{t_1} - \mathbf{z}'_{t_2}, 2\sigma^2\right) \tag{12}$$

$$\int f_2^2(\mathbf{z})\,d\mathbf{z} = \frac{\left(\sqrt{2\pi}\sqrt{2}\sigma\right)^{-d'}}{N^2}\sum_{t_1=1}^{N}\sum_{t_2=1}^{N}G\left(\mathbf{z}''_{t_1} - \mathbf{z}''_{t_2}, 2\sigma^2\right) \tag{13}$$

$$\int f_1(\mathbf{z})\,f_2(\mathbf{z})\,d\mathbf{z} = \frac{\left(\sqrt{2\pi}\sqrt{2}\sigma\right)^{-d'}}{N^2}\sum_{t_1=1}^{N}\sum_{t_2=1}^{N}G\left(\mathbf{z}'_{t_1} - \mathbf{z}''_{t_2}, 2\sigma^2\right) \tag{14}$$

Thus, submitting **Eqs.12**,**13**, and **14** into **Eq.11**, we can immediately derive the novel objective function as follows

$$\begin{aligned} E_2 = \frac{\left(\sqrt{2\pi}\sqrt{2}\sigma\right)^{-d'}}{N^2}\Bigg[&\sum_{t_1=1}^{N}\sum_{t_2=1}^{N}G\left(\mathbf{z}'_{t_1} - \mathbf{z}'_{t_2}, 2\sigma^2\right) \\ &+ \sum_{t_1=1}^{N}\sum_{t_2=1}^{N}G\left(\mathbf{z}''_{t_1} - \mathbf{z}''_{t_2}, 2\sigma^2\right) - 2\sum_{t_1=1}^{N}\sum_{t_2=1}^{N}G\left(\mathbf{z}'_{t_1} - \mathbf{z}''_{t_2}, 2\sigma^2\right)\Bigg] \end{aligned} \tag{15}$$

Since **Eq.15** actually originates from the Parzen window desity estimator and relative entropy for the sampling set and roots at the whole distribution of the training sample set, this novel objective function has the following virtues: since the new criterion is based on the density probability and not the local data points, this corresponding model parameter learning can effectively avoid the over-fitting drawback and show a less sensitivity to noise in the noisy environment. Our experimental results in this study will confirm these virtues.

3.3. Entropy-Criterion Based Parameter Learning

For a given modeling model, with the commonly used gradient descent procedure [18], we can easily get the following model parameter's learning rule,

$$p(t+1) = p(t) - r\frac{\partial E_2}{\partial p} \tag{16}$$

where p denotes the model parameter; t denotes the iteration number and r is the learning rate.

4. RESULTS

In this section, we will illustrate the performance of the proposed entropy-criterion based fermentation process modeling on EPS production from Lactobacillus.

4.1. Performance Index

In order to do the comparative study for the performances of different modeling methods with MSE-criterion and entropy-criterion, we adopt the following performance index to evaluate different modeling methods [19,20].

$$J = \sqrt{\frac{\sum_{l=1}^{N}\left(y_l - y_l^d\right)^2}{\sum_{l=1}^{N}\left(y_l^d - \bar{y}\right)^2}} \tag{17}$$

where $\bar{y} = \frac{1}{N}\sum_{l=1}^{N} y_l^d$; N denotes the number of the testing samples; y_l^d is the l th desired output in the testing set; y_l is the predicted output of the model in testing set. Here, the smaller the value of J is, the better the performance of the corresponding training model is.

4.2. Results

In our experiments, we take three modeling methods: MLP network model, RBF network model and TSK fuzzy system model. All three models have four input nodes representing the four influential process variables (concentrations of lactose, casein hydrolysate and triammonium citrate, and inoculum size) and one output node representing the EPS yield (g/l) at the end of batch. The process data for modeling are generated by carrying out a number of fermentation runs under various input conditions. Here we collect 54 sample data as shown in **Table 1**, each sample data represents a pair of model inputs (fermentation conditions) and a single output (EPS concentration). For MLP network model, RBF neural network model and TSK fuzzy system model, these 54 sample data will be partitioned into a training set (45 samples) and a testing set (9 samples) [7]. The training set is utilized to adjust the parameters of all three models and the testing set is used to evaluate the prediction accuracy. The EPS yield comparisons of the sample data and predicted ones in the testing set obtained by using MSE-criterion based models and entropy-criterion based models are illustrated in **Figures 2-4**. a L, T, C and I in the table represented for Lactose/(g/l), Triammonium citrate/(g/l), Casein hydrolysate/(g/l), Inoculum size/(vol%), respectively.

In fact, due to the extremely complexity of both the fermentation mechanism and the limitation of the experimental condition, experimental data may inevitably contain noise. Hence, how to enhance robustness of the fermentation process modeling is very important. In order to compare the robustness between MSE-criterion based models and entropy-criterion based models, we add Gaussian white noise (G(0, σ_1)) to the training sample set, where $\sigma_1 \in (0, 0.20)$ [8]. In **Tables 2-4**, we list the corresponding performance index for the testing set with 11 different Gaussian white noises.

4.3. MLP Network Modeling

Multi-Layer Perceptron (MLP) network [21] is one of the most widely utilized paradigms in the fermentation process modeling, because it is very simple, general and matured. In the network training procedure, the tangent sigmoid activation function and linear combination function are used for computing the outputs of the hidden and output nodes, respectively. When developing an appropriate MLP model, we must carefully select the number of hidden nodes and then use Back- propagation procedure (BP procedure) [22] to adjust the model parameters. Here the MLP network model contains 15 hidden nodes, and its architecture is illustrated in **Figure 1(a)**. The experimental results about EPS fermentation data from Lactobacillus are illustrated in **Table 2**.

4.4. RBF Network Modeling

Another widely utilized modeling method is Radial Basis Function (RBF) neural network [23], Just like MLP network, RBF network is essentially a feed-forward network. However, RBF network utilizes radial basis functions as its activation functions in the hidden layer. In our experiments, the number of hidden nodes is fixed to be 13, and the RBF network's architecture can be seen in **Figure 1(b)**. The experimental results about EPS fermentation data are illustrated in **Table 3**.

4.5. TSK Fuzzy System Modeling

Takagi-Sugeno-Kang (TSK) fuzzy system [24] has been widely applied, due to its strong capability in learning, universal approximation and handling with matural linguistics with fuzzy rules acqruied from the skilled worker and/or experts. In our experiments, the number of the fuzzy rules is fixed to be 8, and the architecture of the TSK fuzzy system can be seen in **Figure 1(c)**. The experimental results about EPS fermentation data are illustrated in **Table 4**.

As it can be seen from **Tables 2**, **3** and **4**, the prediction accuracies of these three modeling methods with the proposed entropy-criterion based objective function are obviously higher than these methods with MSE-criterion based objective function. This fact means that the proposed objective function is very suitable for the EPS fermentation process modeling.

Figures 5-7 are generated from **Tables 2-4**. In **Figures 5-7**, X-axis denotes the added noise corresponded (see the second column in **Tables 2-4**), and Y-axis denotes the testing performance index. Dotted lines correspond to the testing performance indices of MSE-based criterion (see the third column in **Tables 2-4**), while real lines correspond to the testing performance indices of these modeling methods with entropy-based criterion (see the fourth column in **Tables 2-4**).

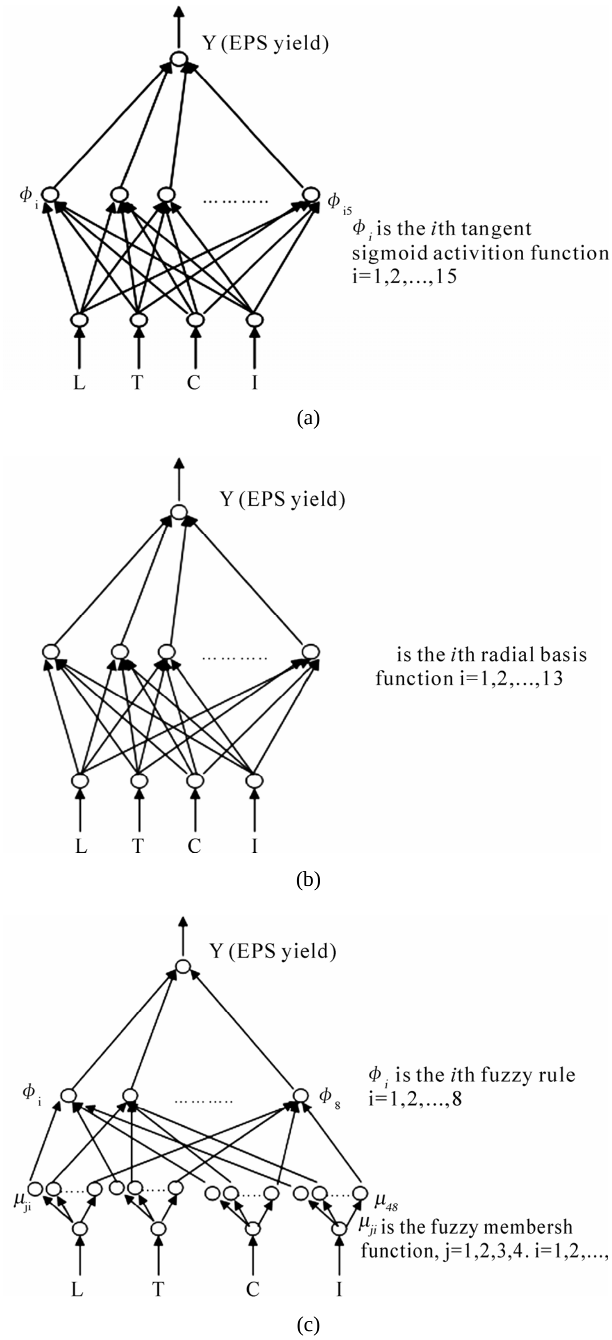

Figure 1. (a) Architecture of MLP network; (b) Architecture of RBF neural $\mu_n$etwork; (c) Architecture of TSK fuzzy system.

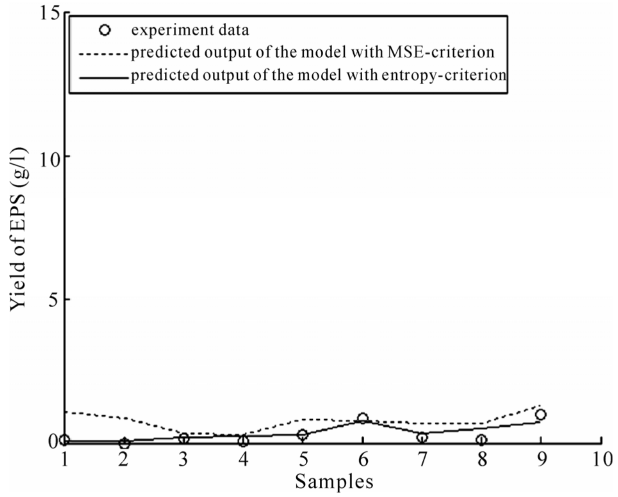

Figure 2. Comparison of EPS yield prediction using MLP network model.

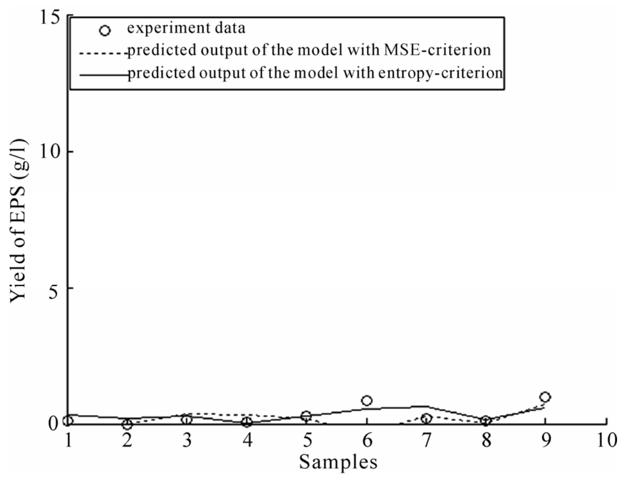

Figure 3. Comparison of EPS yield prediction using RBF model.

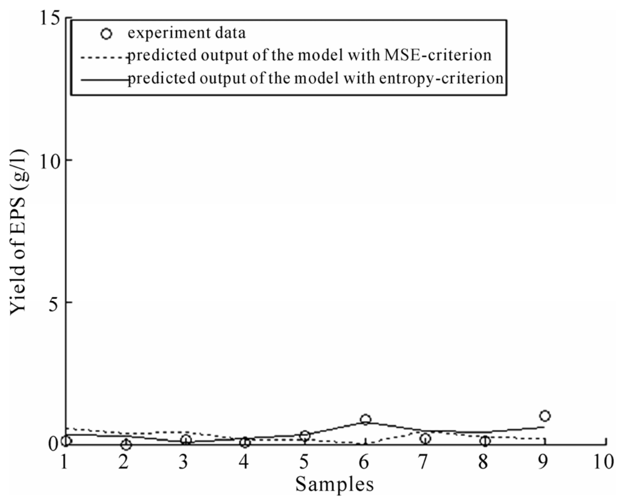

Figure 4. Comparison of EPS yield prediction using fuzzy system model.

Table 1. EPS fermentation data.

No.	Factors and levels				EPS	No.	Factors and levels				EPS
	L	T	C	I			L	T	C	I	
1	8	0	2	1.5	2.29 ± 0.41	28	20	0.3	5	1	1.90 ± 0.44
2	8	0.2	8	3.5	3.43 ± 0.12	29	8	0.2	4	3.5	2.68 ± 0.42
3	25	0	4	1.5	5.16 ± 0.73	30	8	0	4	2	2.65 ± 0.45
4	8	0.2	2	2	2.38 ± 0.62	31	8	0.2	2	1.5	2.51 ± 0.66
5	25	0.2	2	2	4.51 ± 0.37	32	4	0.2	8	2	1.99 ± 0.07
6	25	0.2	8	2	5.32 ± 0.12	33	25	0	4	3.5	5.04 ± 0.16
7	20	0.2	8	1	1.64 ± 0.18	34	40	0.2	4	1	1.88 ± 0.05
8	4	0.2	4	1.5	1.54 ± 0.04	35	2	0.2	4	1	0.65 ± 0.46
9	25	0.2	4	2	5.20 ± 0.24	36	8	0.2	8	2	3.55 ± 0.40
10	25	0	8	1.5	5.66 ± 0.76	37	8	0.2	4	1	1.20 ± 0.05
11	8	0	4	1.5	2.80 ± 0.01	38	20	0.2	4	1	1.70 ± 0.18
12	20	0.1	5	1	1.80 ± 0.50	39	25	0	2	3.5	4.61 ± 0.73
13	10	0.2	4	1	1.35 ± 0.64	40	4	0	8	1.5	2.26 ± 0.48
14	4	0	4	1.5	1.43 ± 0.15	41	4	0.2	8	3.5	2.17 ± 0.39
15	25	0.2	4	1.5	5.22 ± 0.57	42	20	0.2	1	1	0.80 ± 0.69
16	4	0	2	2	0.98 ± 0.58	43	4	0	2	3.5	1.02 ± 0.34
17	8	0.2	4	1.5	2.91 ± 0.32	44	25	0	2	2	4.90 ± 0.57
18	8	0.2	8	1.5	3.79 ± 0.53	45	4	0.2	2	3.5	1.11 ± 0.21
19	4	0.2	2	1.5	1.08 ± 0.42	46	20	0.2	5	1	1.95 ± 0.26
20	4	0.2	4	1	0.80 ± 0.51	47	20	0.2	3	1	1.40 ± 0.13
21	25	0	4	2	5.40 ± 0.12	48	4	0.2	8	1.5	1.98 ± 0.79
22	4	0	4	2	1.59 ± 0.34	49	4	0.2	4	2	1.60 ± 0.73
23	25	0	8	3.5	5.13 ± 0.30	50	4	0.2	4	3.5	2.53 ± 0.28
24	8	0	2	2	2.59 ± 0.59	51	25	0.2	2	3.5	5.04 ± 0.69
25	8	0	4	3.5	2.87 ± 0.47	52	4	0	8	3.5	2.25 ± 0.12
26	8	0	8	3.5	3.78 ± 0.52	53	20	0.4	5	1	1.86 ± 0.26
27	4	0	8	2	2.21 ± 0.71	54	25	0	8	2	5.64 ± 0.66

Table 2. The results about MLP network modeling with MSE-criterion and entropy-criterion.

No.	Noise	Performance index J	
		MSE-criterion	Entropy-criterion
1	0.00	0.8621	0.7730
2	0.02	0.9101	0.8160
3	0.04	0.9940	0.7988
4	0.06	1.1855	0.8487
5	0.08	1.0616	0.8502
6	0.10	1.2130	0.8501
7	0.12	1.3540	0.8511
8	0.14	1.4107	0.8565
9	0.16	1.3558	0.8676
10	0.18	1.5425	0.9117
11	0.20	1.5924	0.9188

Table 3. The results about RBF network modeling with MSE-criterion and entropy-criterion.

No.	Noise	Performance index J	
		MSE-criterion	Entropy-criterion
1	0.00	0.8724	0.7223
2	0.02	0.9279	0.7291
3	0.04	0.9375	0.7441
4	0.06	0.9637	0.7572
5	0.08	0.9970	0.7522
6	0.10	0.9718	0.7779
7	0.12	1.0933	0.7753
8	0.14	1.1369	0.7868
9	0.16	1.1793	0.7914
10	0.18	1.2189	0.7999
11	0.20	1.2413	0.8051

Table 4. The results about fuzzy system modeling with MSE-criterion and entropy-criterion.

No.	Noise	Performance index J	
		MSE-criterion	Entropy-criterion
1	0.00	0.8810	0.6889
2	0.02	1.0852	0.7446
3	0.04	1.1616	0.7641
4	0.06	1.2195	0.8051
5	0.08	1.2237	0.8170
6	0.10	1.2531	0.8124
7	0.12	1.2330	0.8235
8	0.14	1.2689	0.8201
9	0.16	1.2680	0.8390
10	0.18	1.2798	0.8771
11	0.20	1.3141	0.8733

From **Figures 5-7**, it is easy to observe that the three curves corresponding to these three modeling methods with MSE-criterion objective function are always respectively over the curves of these three modeling methods with entropy-criterion based objective function. In addition, with the increases of the noise, the curves of predicted performance indices in **Figures 5-7**, corresponding to the MSE-criterion based modeling methods, have dramatic changes, which mean that the prediction accuracy is deteriorated greatly with the increasing of noise, while the curves corresponding to entropy-criterion based modeling methods in these figures are very smooth. Therefore the experimental results obviously demonstrate that the entropy-criterion based modeling

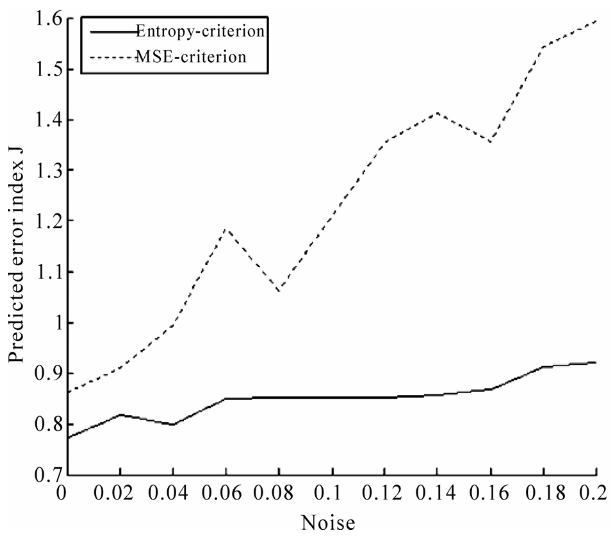

Figure 5. Comparison of testing performance indices of MSE-criterion and entropy-criterion based MLP network modeling method.

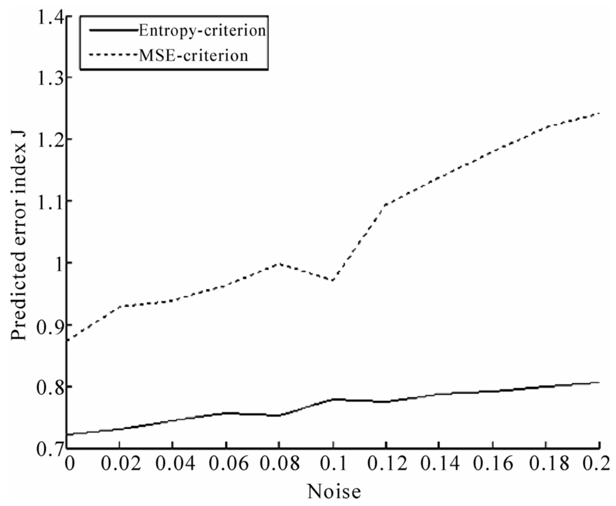

Figure 6. Comparison of testing performance indices of MSE-criterion and entropy-criterion based RBF modeling method.

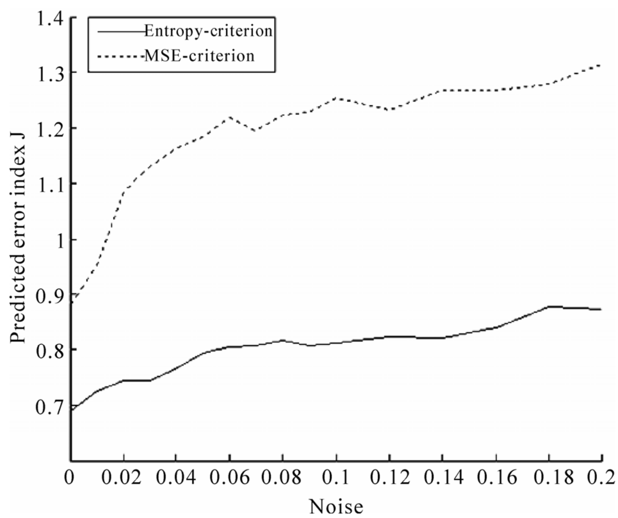

Figure 7. Comparison of testing performance indices of MSE-criterion and entropy-criterion based fuzzy system modeling method.

methods have a better generalization and robustness than the MSE-criterion based modeling methods in the EPS fermentation process modeling.

4.6. Statistical Results for the Obtained Performance Indices

In view of the mean and standard variance of EPS production obtained from the above experiments as the output of the training samples, we can see from **Table 1** that the standard variance is not little, therefore, it is necessary for us to observe the performance of the above three modeling methods from the statistical viewpoint.

In this experiment, we keep the same inputs in the training set as above, however, add noise to the corresponding outputs. The added noise has the mean zero and the same standard variance as derived from the experimental data. In order to keep the experimental results fair, we run each sample data 50 times, and then take their means and standard variances of the performance indices J for the corresponding modeling methods. **Table 5** lists the obtained results.

We can clearly see from **Table 5** that, both the means and standard variances of the outputs of these three modeling methods with entropy-criterion are always lower than the ones with MSE-criterion. This fact confirms our claims again that the proposed entropy- criterion based modeling methods possess the favorable capability in approximation, generalization and robustness.

5. DISCUSSION

When studying fermentation process modeling of EPS from Lactobacillus, we must consider two factors. One is the collected data corrupted by noise, due to the shortage of apparatus and the limitation of experimental conditions. The other is the comparatively weak generalization and robustness capability of current MSE-criterion based modeling methods. In this work, the EPS fermentation process modeling methods with entropy-criterion based objective function are addressed. When it is used in MLP network modeling, RBF modeling and TSK fuzzy system modeling for EPS fermentation from Lactobacillus, our experimental results demonstrate that three modeling methods with entropy-criterion are less sensitive to noise and have better generalization abilities and robustnesses than three modeling methods with MSE-criterion. Because the proposed objective function is derived from the Parzen window desity estimator and relative entropy, and considers the whole distribution structure of the training set in the parameter's learning process, which is different from previous study. The results obtained in this study are very useful in modeling EPS fermentation process, and the entropy- criterion based modeling methods can also be efficiently applied to other fermentation processes.

Table 5. Statistical results of the performance index J of three modeling methods.

Modeling methods	J of MLP network	J of RBF network	J of TSK fuzzy system
MSE-criterion	1.3663 ± 0.2577	1.3893 ± 0.1684	1.1197 ± 0.1807
Entropy-criterion	1.1165 ± 0.1363	0.9998 ± 0.1119	0.9240 ± 0.1210

6. ACKNOWLEDGEMENTS

This work was supported by 863 project of china (grant No.2007AA1Z158, 2006AA10Z313), National Science Foundation of China (grant No. 60773206/F020106，60704047/F030304), The Project Innovation of Graduate Students of Jiangsu Province of China 2008, New century Outstanding Young Scholar Grant of Ministry of Education of China (NCET-04-0496), 2006 Outstanding Young Scholar Grant at JiangSu Province, Grants from the National KeySoft Lab. at Nanjing University and from Jiangsu Province Key Lab. of Information Technologies at Suzhou university.

REFERENCES

[1] Du, F., Lei, M. and Liu, Q. (2004) Advanced control in fermentation process. *Chinese Journal of Information Technology in Construction*, **33**, 314-317.

[2] Gao, X.J., Wang, P., Sun, C.Z., Zhang, Y.T., Zhang, H.Q. and Fan, Q.W. (2006) Modeling and optimization control for the microbial fermentation process. *Chinese Journal of Control Science and Engineering*, **13**, 152-153.

[3] Lee, M.W., Hong, S.H., Choi, H., Kim, J.H., Lee, D.S. and Park, J.M. (2008) Real-time remote monitoring of small-scaled biological wastewater treatment plants by a multivariate statistical process control and neural network-based software sensors. *Process Biochemistry*, **43**, 1107-1113.

[4] Desai, K.M., Vaidya, B.K., Singhal, R.S. and Bhagwat, S.S. (2005) Use of an artificial neural network in modeling yeast biomass and yield of b-glucan. *Process Biochemistry*, **40**, 1617-26.

[5] Kennedy, M.J., Prapulla, S.G. and Thakur, M.S. (1992) A comparison of neural networks to factorial design. *Biotechnology Techniques*, **6**, 293-299.

[6] Zhang, G.Y. and Fang, B.S. (2005) A model for amino acid composition and optimum pH in G/11 xylanase based on neural networks. *Chinese Journal of Biotechnology*, **21(4)**, 658-661.

[7] Desai, K.M., Akolkar, S.K., Badhe, Y.P., Tambe, S.S. and Lele, S.S. (2006) Optimization of fermentation media for exopolysaccharide production from Lactobacillus plantarum using artificial intelligence based techniques. *Process Biochemistry*, **41**, 1842-1848.

[8] Tan, Z.P., Wang, S.T. and Du, G.C. (2008) Glutathione Fermentation Process Modeling based on CCTSK Fuzzy Neural Network. *Journal of Biotechnology*, **7**, 73-79.

[9] Yin, M., Zhang, X.H. and Dai, X.Z. (2000) Dissolved oxygen predictive control based on fuzzy neural networks for fermentation process. *Chinese Journal of Control and Decide*, **15**, 523-526.

[10] Feng, B. and Xu, W.B. (2006) Biochemical variable estimation model based on TSK fuzzy system. *Chinese Journal of Applied Chemistry*, **23**, 343-346.

[11] Meng, H., Fu, X.M. and Cao, G.P. (1999) Neural network fuzzy control of critric acid fermentation process. *Chinese Hebei Journal of Industrial Science & Technology*, **16**, 53-55.

[12] Bian, Z.Q., Zhang, X.G., Yan, P.F., Zhao, N.Y. and Zhang, C.S. (1999) Pattern recognition. Press of Tsing-hua University, Beijing.

[13] Deng, Z.H. and Wang, S.T. (2004) RBF regression modeling based on visual system theory and weber law.

Journal of Southern Yangtze University (Natural Science Edition), **3**, 25-29.

[14] Wang, S.H., Chung, F.L., Xu, M., Deng, Z.H. and Hu, D.W. (2007) A visual system theoretic cost criterion and its application to clustering. *Information Technology Journal*, **6**, 310-324.

[15] Kullback, S. (1968) Information theory and statistics. Dover Publications, New York.

[16] Girolami, M. and He, C. (2003) Probability density estimation from optimally condensed data samples. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **25**, 1253-1264.

[17] Erhan, G. and Jose, C.P. (2002) Information theoretic clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24**, 158-171.

[18] Gardner, W.A. (1984) Learning characteristics of stochastic-gradient-descent algorithms: A general study, analysis and critique. *Signal Processing*, **6**, 113-133.

[19] Jang, J.S.R., Sun, C.T. and Mizutani, E. (1997) A computational approach to learning and machine intelligence. *Neuro-Fuzzy and Soft Computing*, Upper Saddle River, NJ: Prentice-Hall.

[20] Chung, F.L., Wang, S.T., Deng, Z.H. and Hu, D.W. (2006) CATSMLP: Towards a robust and interpretable multilayer perceptron with sigmoid activation functions. *IEEE Transactions on Systems, Man, and Cybernetics-Part B*, **36**, 1319-1331.

[21] Arena, P.L., Fortuna, R.R. and Xibilia, M.G. (1995) Multilayer perceptrons to approximate comples valued functions. *International Journal of Neural Systems*, **6**, 435-446.

[22] Chen, D.S. and Jain, R.C. (1994) A robust backpropagation learning algorithm for function approximation. *IEEE Transactions on Neural Networks*, **5**, 467-479.

[23] Park, J. and Sandberg, I.W. (1991) Universal approximation using radial-basis-function networks. *Neural Computation*, **3**, 246-257.

[24] Sugeno, M. and Yasukawa, T. (1993) A fuzzy-logic-based approach to qualitative modeling. *IEEE Transactions on Fuzzy Systems*, **1**, 7-31.

