Sensors & Transducers

Soft Sensing of Key State Variables in Fermentation Process Based on Relevance Vector Machine with Hybrid Kernel Function

Xianglin ZHU, Xiaofu JI
School of Electrical and Information Engineering, Jiangsu University, 212013, P. R. China

Received: 17 May 2014 /Accepted: 16 June 2014 /Published: 30 June 2014

Abstract: To resolve the online detection difficulty of some important state variables in fermentation process with traditional instruments, a soft sensing modeling method based on relevance vector machine (RVM) with a hybrid kernel function is presented. Based on the characteristic analysis of two commonly-used kernel functions, that is, local Gaussian kernel function and global polynomial kernel function, a hybrid kernel function combing merits of Gaussian kernel function and polynomial kernel function is constructed. To design optimal parameters of this kernel function, the particle swarm optimization (PSO) algorithm is applied. The proposed modeling method is used to predict the value of cell concentration in the Lysine fermentation process. Simulation results show that the presented hybrid-kernel RVM model has a better accuracy and performance than the single kernel RVM model. *Copyright © 2014 IFSA Publishing, S. L.*

Keywords: Hybrid kernel function, Relevance vector machine (RVM), Particle swarm optimization (PSO), Fermentation procedure, Soft sensor.

1. Introduction

The fermentation process is a very complex bio-chemistry procedure that is characterized by multi-variable, system coupling, and strong nonlinearity, since some complicated biological, thermodynamic and physical reactions are involved simultaneously. Especially, the variable of cell concentration mirroring the fermentation procedure quality is difficult to detect online, which makes it difficult to apply online optimal control strategy. The traditional method to obtain this variable is online sample and offline analysis which is characterized by time-consumption and delayed measurement. For this reason, the traditional method did not meet the requirement of online control. Recently, a newly developed method called soft sensing can efficiently resolve this problem [1, 2]. The idea of soft sensing originated from the inferential control theory where primary variables difficult to detect online is numerically estimated by some assistant variables easy to detect through measuring instruments [1, 3].

The commonly used soft sensing methods are based of the artificial neural network (ANN) and support vector machine (SVM) respectively. For examples, Gonzaga *et. al.* used ANN to measure key process variables online in polyethylene terephthalate (PET) manufacturing processes [2], and Desai *et. al.* built the soft sensing model of fed-batch fermentation procedures [4], and Liu *et. al.* gave a soft sensing model in erythromycin fermentation processes [5]. The relevance vector machine (RVM) is a recently-developed learning algorithm based on the sparse Bayesian learning theory, whose structure is similar to SVM [6]. However, unlike the SVM, the RVM model has some excellent merits such as high sparsity, good prediction precision, strong generalization ability, ability to give probabilistic model output, and so on [6, 7].

The kernel function is an important element of the RVM that vitally affects the RVM performance. Unlike kernel functions of SVM, kernel functions of RVM are not required to satisfy the Mercer condition, which makes the construction of RVM's kernel functions more flexible than that of SVM. The frequently applied kernel functions of RVM are local kernel functions presented by the Gaussian kernel function and the global kernel functions presented by the polynomial kernel function. For these two classes of kernel functions, the local kernel function has strong local learning ability and poor generalization ability, while the global kernel function is with poor learning stability and strong generalization ability.

To resolve the online measurement problem of important state variables in fermentation processes with traditional instruments, a novel soft sensing modeling method based on RVM with a hybrid kernel function is presented in this paper. Combining merits of the local Gaussian kernel function and global polynomial kernel function, a novel hybrid kernel function is constructed, whose key kernel parameters are optimally designed using the particle swarm optimization (PSO) algorithm. To resolve the online detection of cell concentration in the Lysine fermentation process, the proposed modeling method is applied to build soft sensing model. The simulation and experiment show that the designed RVM model with hybrid kernel function achieves a better fitting performance and generalization ability than RVM models with Gaussian kernel function and polynomial kernel function respectively.

The rest of this paper is organized as follows. Section 2 gives some fundamentals of RVM and the general RVM model procedure. Section 3 describes typical characteristics of local kernel function and global kernel function, based on which, the basic idea of hybrid kernel function is given in Section 4. In the section, PSO algorithm is also applied to design the optimal kernel parameters. In section 5, the proposed RVM soft sensing method is used to predict the value of cell concentration in the Lysine fermentation process, whose result verifies the effectiveness of the presented method. Section 6 gives the condition of this paper.

2. Relevance Vector Machine

For a given training set $\{x,t\}_{i=1}^{n}$, the objective regression function t of RVM is supposed to be a nonlinear function contaminated by a white Gaussian noise,

$$t = y(x,\omega) + \varepsilon, \tag{1}$$

where the noise ε follows the Gaussian distribution with mean 0 and variance σ^2, that is, $p(\varepsilon \mid \sigma^2) \sim \mathcal{N}(0,\sigma^2)$. By following the same philosophy as that in the derivation of support vector machines, RVM model expresses the regression function $y(x,\omega)$ as a combination of kernel function $K(x,x_i)$,

$$y(x,\omega) = \sum_{i=1}^{n} \omega_i K(x,x_i) + w_0 \tag{2}$$

where $\omega = [\omega_0, \omega_1, \ldots, \omega_n]^T$ is the weighing vector, and then the likelihood function of training samples can be given as

$$\begin{aligned} y(x,\omega) &= \prod_{i=1}^{n} \mathcal{N}(t_i \mid y(x_i,\omega), \sigma^2) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \| t - \Phi\omega\|^2\right) \end{aligned} \tag{3}$$

where $t = [t_1, t_2, \ldots, t_n]^T$, and

$$\Phi = [\phi(x_1), \phi(x_2), \ldots, \phi(x_n)] = \begin{bmatrix} 1 & K(x_1,x_1) & K(x_1,x_2) & \cdots & K(x_1,x_n) \\ 1 & K(x_2,x_1) & K(x_2,x_2) & \cdots & K(x_2,x_n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & K(x_n,x_1) & K(x_n,x_2) & \cdots & K(x_n,x_n) \end{bmatrix}$$

The prior distribution of weighing parameters ω_i, $(i = 0,1,\ldots,n)$ is assumed to be Gaussian ones dependent on the hyper-parameter α, that is

$$p(\omega \mid \alpha) = \prod_{i=0}^{n} \mathcal{N}(\omega_i \mid 0, \alpha^{-1}) \tag{4}$$

where $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_n]^T$ is the hyper-parameter vector determining the prior distribution of weighing vector ω, and further determining the sparsity performance of the obtained RVM model. According to the Bayesian rule, the posterior distribution of ω can be given as,

$$\begin{aligned} p(w \mid t,\alpha,\sigma^2) &= \frac{p(t \mid w,\sigma^2) p(w \mid \alpha)}{p(t \mid \alpha,\sigma^2)} \\ &= (2\pi)^{-\frac{n}{2}} \left|\Sigma\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(w-\mu)^T \Sigma^{-1}(\omega-\mu)\right), \end{aligned} \tag{5}$$

where the posterior covariance of this distribution is $\Sigma = (\sigma^{-1}\Phi^T\Phi + A)^{-1}$, the posterior mean is $\mu = \sigma^2\Sigma\Phi^T t$, and A is a diagonal matrix with $A = \operatorname{diag}(\alpha_0, \alpha_1, \ldots, \alpha_n)$. By integrating the likelihood function (3) with weighing vector, we can obtain the edge distribution of t dependent on α and σ^2 as

$$p(t,|\alpha,\sigma^2) = (2\pi)^{-\frac{n}{2}} |\Omega|^{-\frac{1}{2}} \exp\left(-\frac{t^T\Omega t}{2}\right), \quad (6)$$

where $\Omega = \sigma^2 I + \Phi A^{-1}\Phi^T$.

To obtain optimal parameters α and σ^2 with maximizing (6), we can calculate the partial derivation of (6) and apply the iteration method. With this philosophy, the updating equation of hyper-parameters α and σ^2 can be given as

$$\alpha_i^{\text{new}} = \frac{\gamma_i}{\mu_i} \quad (7)$$

$$(\sigma^2)^{\text{new}} = \frac{\| t - \Phi\mu\|^2}{\sum_{i=1}^{n}\gamma_i}, \quad (8)$$

where $\gamma_i = 1 - \alpha_i \Sigma_{ii}$, μ_i is the ith posterior mean, Σ_{ii} is the ith diagonal element of the posterior covariance matrix Σ. It can be seen that $\gamma_i \in [0,1]$ can measure the fitness performance of weigh ω_i to training samples. Through the iteration (7), (8), μ and Σ are updated accordingly. In this procedure, most α_i, $(i = 1,2,\ldots,n)$ will approach to 0 and the corresponding basis function $\phi(x_i)$ will be deleted, in which way, the high sparsity of RVM is obtained.

For a new input x_*, the predicated distribution of corresponding output t_* is given as

$$\begin{aligned} &(t_* \mid t, \alpha_{\text{MP}}, \sigma^2_{\text{MP}}) \\ &= \int p(t_* \mid w, \sigma^2_{\text{MP}}) p(\omega \mid t, \alpha_{\text{MP}}, \sigma^2_{\text{MP}}) d\omega \\ &\sim \mathcal{N}(t_* \mid y_*, \sigma^2_*) \end{aligned} \quad (9)$$

3. Local and Global Characteristics of Typical Single Kernel Functions

3.1. Local Gaussian Kernel Function

The Gaussian kernel function

$$K(x, x_i) = \exp\left(\frac{-\| x - x_i\|^2}{\sigma^2}\right) \quad (10)$$

is a typical local kernel function [7]. To verify the local characteristic of this function, we choose the testing point $x = 2$ and plot the characteristic curves with different σ^2 as shown in Fig. 1.

It can be shown that $K(x, x_i)$ is comparatively big when the input value x_i is close to $x = 0.2$, and however, this value goes downhill quickly with the distance between x and x_i increasing. Especially, this value approaches to 0 when the sample point goes aloof. This phenomena verifies that the Gaussian kernel function has strong local interpolation ability and comparatively poor generalization ability, and that the performance characteristic of Gaussian kernel function is closely related to σ^2.

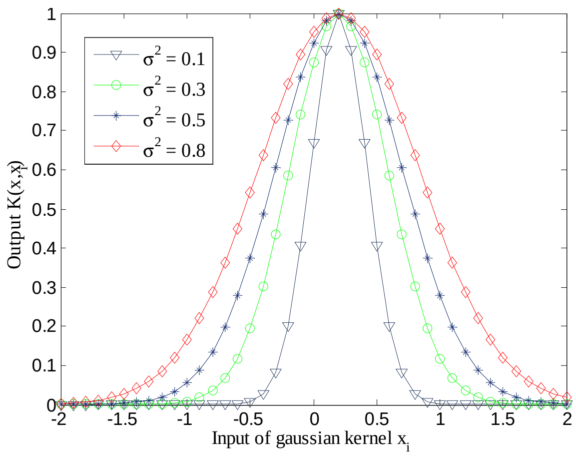

Fig. 1. Characteristic curves of Gaussian kernel function.

3.2. Global Polynomial Kernel Function

The polynomial kernel function

$$K(x, x_i) = \left(\frac{x \cdot x_i}{\sigma^2} + 1\right)^q \quad (11)$$

is the typical global kernel function [8], where the order q determines the global generalization ability. To obtain comparison result with the characteristic of Gaussian kernel function (10), we give the characterristic curves with testing point $x = 0.2$ and different q, σ^2 in Fig. 2.

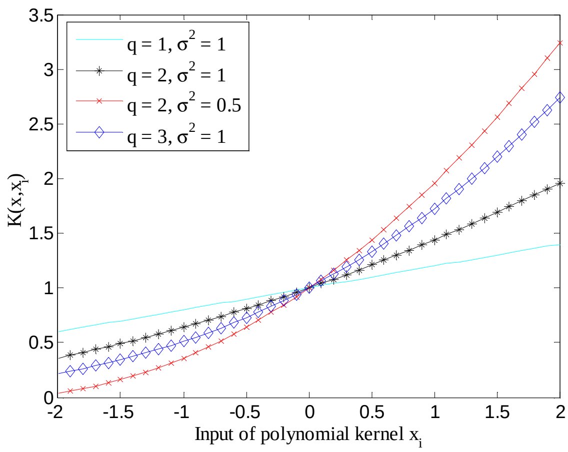

Fig. 2. Characteristic curves of polynomial kernel function.

It can be shown that this function value change gently with the disturbance between the input value x_i and test value x increasing. And thus, the polynomial kernel function has good generalization ability. It can be also shown that the parameter σ^2 determines the affection speed of this kernel function to sample value, and thus the parameter σ^2 can be used to enhance the local interpolation ability of (11) without increasing q.

4. Hybrid Kernel Function and PSO-Based Parameter Design

4.1. Hybrid Kernel Function Design

From above discussion, we can show that the Gaussian kernel function has a good local learning ability and comparatively poor generalization, and that the polynomial kernel function has a strong generalization ability and comparatively poor learning ability. If a hybrid kernel function can be designed combining merits of these two functions, good abilities of local learning and generalization can be achieved.

The hybrid kernel function of this paper is designed as follows

$$K(x,x_i) = (1-\alpha)\left(\frac{x\cdot x_i}{\sigma_p^2}+r\right)^q + \alpha\exp\left(\frac{\| x-x_i\|^2}{\sigma_g^2}\right) \quad (12)$$

where $\alpha \in [0,1]$ is the weighing scalar. It can be easily seen that this kernel function reduces to the traditional polynomial kernel function when $\alpha = 0$, and that it reduces to the Gaussian kernel function when $\alpha = 1$. In this sense, this kernel function is more general than traditional ones and therefore can achieve better performance. In hybrid kernel function (12), the kernel width σ_g^2 can adjust the local fitting ability and generalization ability of Gaussian function, and the order q can adjust the computational complexity and weighting freedom of RVM model.

To analyze the characteristic of this function, we also compares it with the Gaussian kernel function (10) and polynomial kernel function (11) respectively. By choosing the testing point $x = 0.2$, the characteristic curves are shown in Fig. 3, which implies that this hybrid kernel function combines both merits of the Gaussian kernel function and the polynomial kernel function, that is, it achieves comparatively strong learning and generalization ability.

It can be shown from Fig. 3 that curves (2) and (3) imply the influence of parameter r to the characteristic of hybrid kernel function, and that curves (4) and (6) imply the global adjustment of q to hybrid kernel function, and that curves (4) and (5) imply the influence of weighing scalar α to hybrid kernel function. It can be also seen from Fig. 3 that the kernel parameters α, σ^2, r, σ_g, and q are very important to determine the fitting and generalization ability of RVM model.

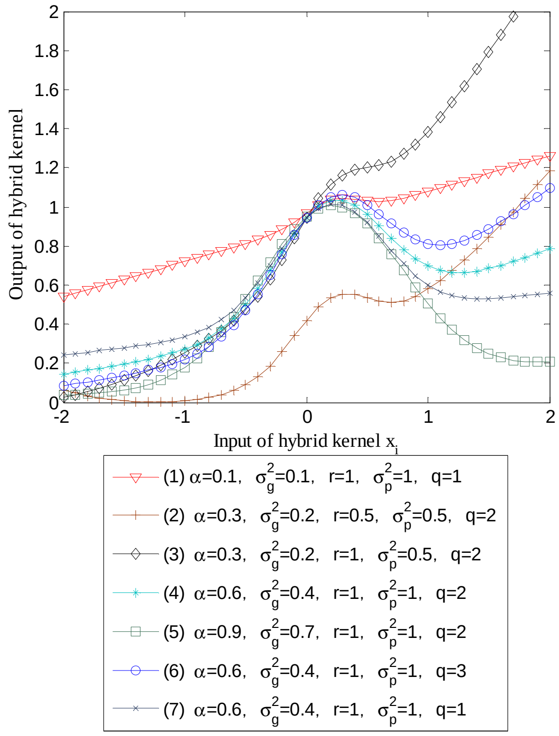

Fig. 3. Characteristic curves of hybrid kernel function.

4.2. Parameter Design of Hybrid Kernel Function using PSO Algorithm

It can be seen from section 4.1 that a strong global generalization ability and local interpolation ability can be achieved if parameters α, σ^2, r, σ_g, and q re suitably designed. However, it is not an easy work since the relationship between these parameters is somewhat complex, which is impossible to be expressed explicitly. With this consideration, PSO algorithm is applied to design optimal parameters, which may comparatively improve the RVM model performance.

PSO algorithm, which is known as a high-quality intelligent evolutionary computation algorithm, was first presented in 1995 [9]. Comparing with the traditional mesh search method and gradient descent method, PSO algorithm is characterized by rapid searching rate, global convergence and high stability. PSO algorithm simulates the social behavior of individuals (particles) “flying” in a multidimensional search space. Each particle in the swarm looks for its optimal value in the search space by utilizing both its own experience and its neighbors’ experience, in which way, PSO algorithm can solve some different optimization problems.

In a D-dimensioned complex search space, the ith particle updates its position and velocity using the following equations,

$$\begin{aligned} v_{id}^{k+1} &= w(t)v_{id}^{k} + c_1 r_1 (p_{\text{best}}^{k} - x_{id}^{k}) + c_2 r_2 (g_{\text{best}}^{k} - x_{id}^{k}) \\ x_{id}^{k+1} &= x_{id}^{k} + v_{id}^{k+1} \end{aligned} \quad (13)$$

where $v_i = (v_{i1}, v_{i2}, \dots, v_{iD})$ and $x_i = (x_{i1}, x_{i2}, \dots, x_{iD})$ are the speed and position vector of the ith particle, k is the iteration step index, r_1 and r_2 are random variables between 0 and 1, c_1 and c_2 are learning factors, p_{best} is the best solution among the particles found in the current iteration, g_{best} is the global beat solution achieved so far, $w(t)$ is the inertia weight which can be described as

$$w(t) = w_{\text{in}} - \frac{w_{\text{in}} - w_{\text{end}}}{T_{\max}} \times t \quad (14)$$

where w_{in} is the original inertia weight, w_{end} is the final inertia weight, $T_{\max}$ is the whole iterative time, t is the current iterative time. It can be shown that $w(t)$ will linearly decrease from w_{in} to w_{end} with iteration proceeding.

The speed vector v_{id} in (13) is constrained by

$$v_{id}^{k+1} = \begin{cases} v_{\max}, & v_{id}^{k+1} \geq v_{\max} \\ -v_{\max}, & v_{id}^{k+1} \leq -v_{\max} \end{cases} \quad (15)$$

The detailed procedure of designing parameters of hybrid kernel function is shown in Fig. 4.

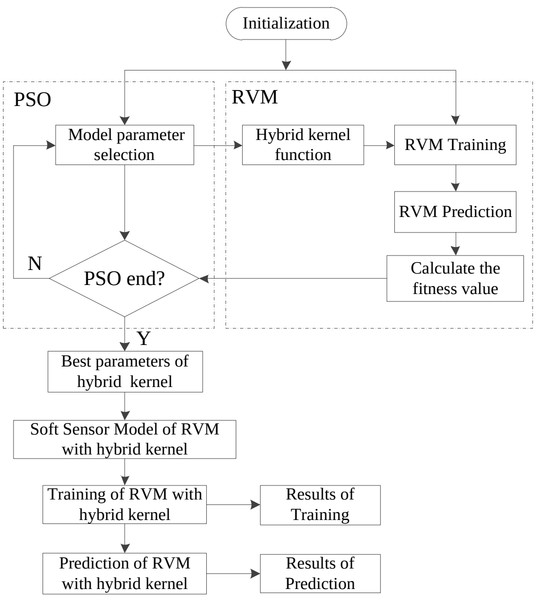

Fig. 4. The flow chart of soft sensor modeling based on RVM of hybrid kernel.

5. Experiments and Simulation of Lysine Fermentation

In this section, the hybrid-kernelled RVM model is built to predict the cell concentration in the Lysine fermentation process. The fermentation experiments are carried out based on WKT-30L fermentation equipment shown in Fig. 5 and the corresponding digital automatic control system.

Fig. 5. WKT-30L fermentation experiment system.

According to the environmental requirement of the Lysine fermentation process, the pressure in fermentation procedure is controlled to be 0.11 MPa, temperature $31^{\circ}\mathrm{C}$, rotating speed of the mixing motor 220 r/min. According to our early research, the supplementary variables are chosen as the variables of pH level (pH), dissolved oxygen (Do), air flux F and the primary variable is chosen as cell concentration y. In this way, the soft sensor model can be designed as

$$y = f(\text{pH}, \text{Do}, F) \quad (16)$$

where $f(\cdot)$ denote the complex nonlinear relationship among these variables.

To obtain enough testing samples, WKT-30L experiment system samples pH, Do, and F every minute. At the same time, the reaction liquid is sampled every 4 hours that gives exact cell concentration via 721-type spectrophotomete. One batch sample can be obtained in a fermentation period. Totally, 5 experiments are carried out and thus 184 samples are obtained, among which, the former 138 samples are used as training data and the rest 6 samples as testing data. To obtain a better model performance, samples are normalized into the interval [0, 1] before RVM training.

The speed constraint in PSO algorithm is chosen as 20 % of $v_{\max}$. This implies, according to (15), that

$w(t)$ will decrease from 0.9 to 0.4 in the iteration procedure. The rest parameters are given as $c_1 = c_2 = 2$, $N = 24$, $T_{max} = 1000$. Following the design flow shown in Fig. 4, the optimal parameters of hybrid kernel function are obtained as $\alpha = 0.4762$, $\sigma_g^2 = 60.9969$, $r = 1.8358$, $\sigma_g^2 = 0.8816$, $q = 2$. The RVM models with hybrid kernel function, polynomial kernel function, Gaussian kernel function, respectively are shown in Fig. 6 (In this Figure, the kernel parameters of RVM models with single kernel function are also designed using PSO algorithm), and the corresponding relative error curves are shown in Fig. 7.

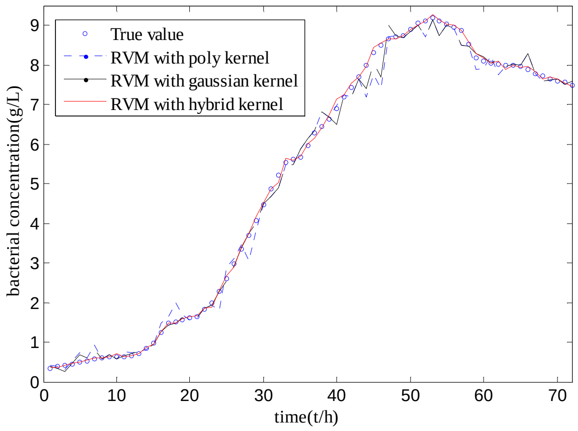

Fig. 6. Comparison of prediction using soft sensor model based on single kernel function and hybrid kernel function.

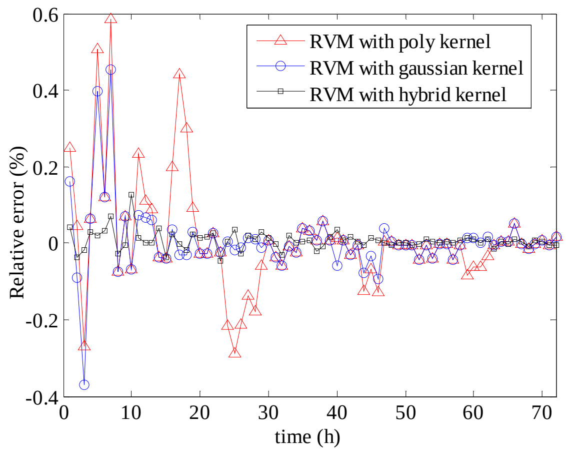

Fig. 7. Comparison of relative error using soft sensor model based on single kernel function and hybrid kernel function RVM.

To further analyze the prediction ability of three models, the following average relative error (ARE), maximum absolute error (MaxE), and root mean squared error (MSE) are defined to evaluate the model performance

$$\text{ARE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - y_i^*}{y_i}\right| \times 100\% \qquad (17)$$

$$\text{MaxE} = \max(|\, y_i - y_i^*\, |) \qquad (18)$$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - y_i^*)^2 \qquad (19)$$

where y_i is the exact value, y_i^* is the predicted value of RVM.

The number of relative vectors (RVS), MSE, ARE and MaxE of three RVM models are given respectively in Table 1. It can be shown that the prediction performance of RVM model with hybrid kernel function is much better than the ones with polynomial kernel function and Gaussian kernel function respectively. Furthermore, the RVM model with hybrid kernel function has less relative vectors than the ones with polynomial kernel function and Gaussian kernel function. This implies that the RVM model with hybrid kernel function is simpler and less time-consumption when it is trained.

Table 1. RVs, MSE, ARE and MaxE of three RVM models.

	RVs	MSE	ARE (%)	MaxE
Gaussian	7	0.0406	8.53	0.7912
Polynomial	7	0.1074	7.57	1.0912
Hybrid	6	0.0043	1.59	0.2366

6. Conclusion

A soft sensing modeling method based on RVM with a hybrid kernel function is presented to resolve the online detection difficulty of some important state variables in fermentation processes with traditional instruments. Based on the characteristic analysis of two traditional frequently-used kernel functions, that is, local Gaussian kernel function and global polynomial kernel function, a hybrid kernel function combing merits of Gaussian function and polynomial function is constructed. PSO algorithm is also applied to obtain optimal parameters of kernel functions. The proposed modeling method is used to predict the cell concentration in the Lysine fermentation process. Simulation results show that the hybrid-kernel RVM model has a better accuracy and performance than single kernel RVM models.

Acknowledgements

This work is supported by the natural science foundation of Jiangsu Province, China under grant BK2011465.

References

[1]. J. Zhang, Z. Xiong, D. Guillaume, A. Lamande, Batch to batch iterative learning control of a fed-batch fermentation process using linearised models, in *Proceedings of the International Conference on Control, Automation, Robotics and Vision (ICARAV'08)*, Melia Hanoi Hanoi, Vietnam, 17-20 December 2008, pp. 745-750.

[2]. J. C. B. Gonzaga, L. A. C. Meleiro, C. Kiang, R. M. Filho, ANN-based soft-sensor for real-time process monitoring and control of an industrial polymerization process, *Computers and Chemical Engineering*, Vol. 5, Issue 1, 2009, pp. 43-49.

[3]. C. Brosillow, Inferential control of process, *American Institute of Chemical Engineers Journal*, Vol. 24, Issue 1, 1978, pp. 485-509.

[4]. K. Desai, Y. Badhe, S. S. Tambe, B. D. Kulkarni, Soft-sensor development for fed-batch bioreactors using support vector regression, *Journal of Biochemical Engineering*, Vol. 27, Issue 3, 2006, 225-239.

[5]. G. Liu, D. Zhou, H. Xu, C. Mei, Model optimization of SVM for a fermentation soft sensor, *Expert Systems with Applications*, Vol. 37, Issue 4, 2010, 2708-2713.

[6]. M. E. Tipping, Sparse Bayesian learning and the relevance vector machine, *Journal of Machine Learning Research*, Vol. 1, 2001, 211-244.

[7]. M. E. Tipping, C. A. Faul, Fast marginal likelihood maximisation for sparse Bayesian models, in *Proceedings of the International Workshop on Artificial Intelligence and Statistics (AISTATS'06)*, Key west, Florida, 3-6 January 2006, pp. 2-6.

[8]. K. R. Müller, S. Mika, G. Rätsch, K. Tsuda, An introduction to Kernel-Based Learning Algorithms, *IEEE Transactions on Neural Networks*, Vol. 12, Issue 2, 2001, pp. 181-202.

[9]. J. Kennedy, R. Eberhart, Particle swarm optimization, in *Proceeding of the International Conference on Neural Network (INNS'95)*, Perth, Western Australia, 1995, pp. 1942-1948.