Chinese J. Chem. Eng., **13** (5) 642 – 648 (2005)

Fuzzy Support Vector Regression Model of 4-CBA Concentration for Industrial PTA Oxidation Process*

ZHANG Ying(张英), SU Hongye(苏宏业), LIU Ruilan(刘瑞兰) and CHU Jian(褚健)**
National Laboratory of Industrial Control Technology, Institute of Advanced Process Control, Zhejiang University, Hangzhou 310027, China

Abstract In the past few years, support vector machines (SVMs) have been applied to many fields, such as pattern recognition and data mining, *etc.* However there still exist some problems to be solved. One of them is that the SVM is very sensitive to outliers or noises because of over-fitting problem. In this paper, a fuzzy support vector regression (FSVR) method is presented to deal with this problem. Strategies based on k nearest neighbor (kNN) and support vector data description (SVDD) are adopted to set the fuzzy membership values of data points in FSVR. The proposed FSVR soft sensor models based on kNN and SVDD are employed to predict the concentration of 4-carboxy-benzaldehyde (4-CBA) in purified terephthalic acid (PTA) oxidation process. Simulation results indicate that the proposed method indeed reduces the effect of outliers and yields higher accuracy.

Keywords purified terephthalic acid, 4-carboxy-benzaldehyde, support vector machines, soft sensor, fuzzy membership

1 INTRODUCTION

Purified terephthalic acid (PTA) is the main raw material for polyester fiber, PET resins, PET films and other products. It is produced by catalytic oxidization of para-xylene (PX). One of the major impurities, 4-carboxybenzaldehyde (4-CBA), is quite difficult to remove by physical means, so the concentration of 4-CBA is regarded as the main index to the quality of PTA product in the PX oxidation process.

The simplified PX oxidation process flowsheet is shown in Fig. 1. Para-xylene, acetic acid solvent, promoter (hydrobromic acid) and catalyst (cobalt acetate, manganese acetate) are continuously metered into feed mixing tank together with the recycling mother liquid. Then the mixed stream is pumped into the reactor, and the air is fed to the reactor through four-inlets. There are two stages in the oxidation reaction process[1], with the first stage being the agitated oxidation reactor, and the second stage being the agitated first crystallizer. Exothermic heat of reaction is removed by condensing the boiling reaction solvent. A portion of this condensate is withdrawn to control the water concentration in the reactor, and the remainder is refluxed back to the reactor[1].

In order to ensure the quality of PTA, the concentration of 4-CBA should be controlled below a certain level. However, in practice the concentration of 4-CBA is analyzed only three times each day by a spectroscopic analyzer[2,3]. As the spectroscopic analysis is a laboratory technique with obvious time delay, the analysis values of the concentration of 4-CBA are not available for real-time control. In order to measure the concentration of 4-CBA, an alternative method is to build a soft sensor model. Using the model, the concentration of 4-CBA can be inferred from the other measured process variables such as temperature, pressure, flow, *etc.* If the concentration of the 4-CBA can be estimated accurately, an advanced control strategy can be applied to directly control its concentration to ensure the quality of the product[4].

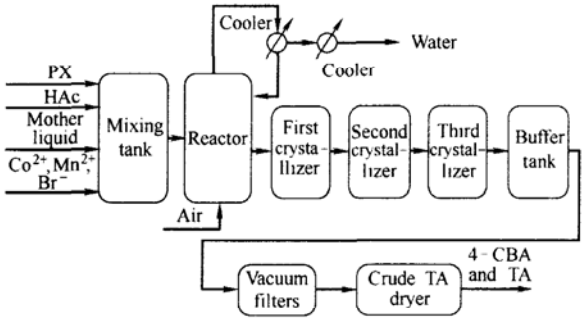


Figure 1 Schematic layout of PX oxidation process

In order to clean the crude TA dryer and adjust the level of buffer tank, load down and up operations have to be performed every week. This process will endure about 8 hours each time. In contrast with the narrow range of 4-CBA concentration in normal operation, the 4-CBA concentration during load down or up operation is very broad[1]. During the load change operation, the data we obtained are different from those collected during normal operation. The former can

Received 2004-10-11, accepted 2005-01-15.
* Supported by National Key Technologies Research and Development Program in the 10th five-year plan (2001BA204B07) and the National Outstanding Youth Science Foundation of China (No. 60025308).
** To whom correspondence should be addressed. E-mail: zhangying@iipc.zju.edu.cn, hysu@iipc.zju.edu.cn

be regarded as outliers. On the one hand, soft sensor model is very sensitive to outliers or noises of training data set, which will affect the accuracy of soft sensor model. On the other hand, in order to predict the concentration of 4-CBA during load down or up operation, the outliers and noise can not be simply removed from the training sample set. In this paper, a fuzzy support vector regression (FSVR) method is introduced to deal with the problem. Strategies of k nearest neighbor (kNN) and support vector data description (SVDD) are adopted to set fuzzy membership values of data points in FSVR. This will significantly reduce the effect of outliers and noises and increase the accuracy of soft sensor model.

2 FUZZY SUPPORT VECTOR REGRESSION

Support vector machines (SVMs) are a new class of machine learning algorithms, motivated by results of statistical learning theory[5]. Because of its well generalization ability and solid theoretical foundation, SVM is applied to many research fields[6]. In order for this property to carry over to the case of regression, Vapnik devised the so-called ε-insensitive loss function and presented SVR algorithm[7]. For a linear separable problem, given the training sample set $\chi = \{(\boldsymbol{x}_1, y_1), \cdots, (\boldsymbol{x}_i, y_i), \cdots, (\boldsymbol{x}_l, y_l)\}$, $i = 1, \cdots, l$, $\boldsymbol{x}_i \in \boldsymbol{R}^n$, $\boldsymbol{y}_i \in \boldsymbol{R}$, the goal of the learning process is to find the smallest risk by minimizing the regularized function as follows[7]:

$$\frac{1}{2}\|\boldsymbol{w}\|^2 + C \cdot R^{\varepsilon}_{\text{emp}}[f] \tag{1}$$

where $\|\boldsymbol{w}\|^2$ is structure risk which characterizes the complexity of model and $R^{\varepsilon}_{\text{emp}}[f]$ is experience risk which represents the training error. Parameter C is slack factor which determines the trade off between structure and experience risk. Minimizing Eq. (1) captures the main insight of statistical learning theory. In order to obtain a small risk, one needs to control both training error and model complexity. For a nonlinear problem, data points in input space are mapped into a high dimension feature space *via* mapping φ: $\boldsymbol{R}^n \rightarrow \boldsymbol{H}$, and then the smallest risk is found in the space $\boldsymbol{H}$. The minimization of Eq. (1) is equivalent to the following quadratic programming (QP) problem

$$\min_{w,\xi,\xi^*} \Phi = \frac{1}{2}\|\boldsymbol{w}\|^2 + C \cdot \sum_{i=1}^{l}(\xi_i + \xi_i^*)$$
$$\text{s.t.} \begin{cases} y_i - (\boldsymbol{w} \cdot \varphi(\boldsymbol{x}_i) + b) \leqslant \varepsilon + \xi_i \\ (\boldsymbol{w} \cdot \varphi(\boldsymbol{x}_i) + b) - y_i \leqslant \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geqslant 0, i = 1, \cdots, l \end{cases} \tag{2}$$

where ξ_i, ξ_i^* are slack variables.

Although SVM is powerful for the problem characterized by small sample, nonlinearity and high dimension[8], there still exist some problems to be solved. One of them is the over-fitting problem. How to set the free parameter is very important in the SVM training process. Recalling the risk function in Eq. (1), a larger C means to assign a higher penalty to error and thus reduces the rate of error. On the contrary, a smaller C is to ignore more plausible error and thus get wider regression margins[9,10]. No matter the value of C is large or small, this parameter is fixed during the training process of SVR. That is to say data points are equally treated during the training process. This will lead to a high sensitivity for some special cases, such as outliers and noises. In order to deal with the problem, fuzzy membership model is introduced to SVR and FSVR[9]. In FSVR, each data point has a membership value μ_i, $0 < \mu_i < 1$. The training data set becomes $(x_1, y_1, \mu_1), \cdots, (x_i, y_i, \mu_i), \cdots, (x_l, y_l, \mu_l)$ and the objection function of optimization question (2) can be rewritten as follows:

$$\min_{\boldsymbol{w},\xi,\xi^*} \Phi = \frac{1}{2}\|\boldsymbol{w}\|^2 + C \cdot \sum_{i=1}^{l} \mu_i(\xi_i + \xi_i^*) \tag{3}$$

Introducing Lagrange multipliers α_i, α_i^*, η_i, η_i^*, we obtain the Lagrangian of Eq. (3)

$$\begin{aligned} L = &\frac{1}{2}\|\boldsymbol{w}\|^2 + C \cdot \sum_{i=1}^{l} \mu_i(\xi_i + \xi_i^*) - \\ &\sum_{i=1}^{l} \alpha_i(\varepsilon + \xi_i - y_i + \boldsymbol{w} \cdot \varphi(\boldsymbol{x}_i) + b) - \\ &\sum_{i=1}^{l} \alpha_i^*(\varepsilon + \xi_i^* + y_i - \boldsymbol{w} \cdot \varphi(\boldsymbol{x}_i) - b) - \\ &\sum_{i=1}^{l}(\eta_i \xi_i + \eta_i^* \xi_i^*) \end{aligned} \tag{4}$$

Let the derivatives with respect to the primal variables equal to zero, we obtain the four equations

$$\begin{cases} \dfrac{\partial L}{\partial \boldsymbol{w}} = \boldsymbol{w} - \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)\varphi(\boldsymbol{x}_i) = 0 \\ \dfrac{\partial L}{\partial b} = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \\ \dfrac{\partial L}{\partial \xi_i} = C\mu_i - (\alpha_i + \eta_i) = 0 \\ \dfrac{\partial L}{\partial \xi_i^*} = C\mu_i - (\alpha_i^* + \eta_i^*) = 0 \end{cases} \tag{5}$$

Substituting the above equation into Eq. (4), Wolfe

dual is obtained as follows:

$$\max W = -\varepsilon\sum_{i=1}^{l}(\alpha_i+\alpha_i^*)+\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)y_i-\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)\varphi(\boldsymbol{x}_i)\cdot\varphi(\boldsymbol{x}_j)$$

$$\text{s.t.}\quad\begin{cases}\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)=0\\ 0\leqslant\alpha_i,\alpha_i^*\leqslant\mu_i C,\quad i=1,\cdots,l\end{cases}\tag{6}$$

According to Mercer condition[8], there exist a mapping φ and a kernel function $K(\cdot,\cdot)$ which satisfies $K(\boldsymbol{x}_i,\boldsymbol{x}_j)=\varphi(\boldsymbol{x}_i)\cdot\varphi(\boldsymbol{x}_j)$. Substituting the kernel function into Eq. (6) and solving the optimization problem, we obtain the regression function as follows:

$$f(\boldsymbol{x})=\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)K(\boldsymbol{x}_i,\boldsymbol{x})+b\tag{7}$$

From Eq. (6) we can see that the upper bound of Lagrange multipliers α_i, α_i^* is a function of fuzzy membership μ_i. Thus this method is called fuzzy support vector regression.

In Table 1, we summarize the common types of kernels. The parameters p and d are user-specified.

Table 1 Three common types of kernels used in SVR

Kernel's types	$K(\boldsymbol{x}_i,\boldsymbol{x}_j)$
linearly kernel	$\boldsymbol{x}_i\cdot\boldsymbol{x}_j$
polynomial kernel	$[(\boldsymbol{x}_i\cdot\boldsymbol{x}_j)+1]^d$
Gauss RBF kernel	$\mathrm{e}^{-p\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2}$

3 FUZZY MEMBERSHIP MODEL

Fuzzy membership model is very important which determines the performance of FSVR. For the sequential learning problem, Lin proposed a membership model[9]

$$\mu_i=f(t_i)=\frac{1-\sigma}{t_l-t_1}t_i+\frac{t_l\sigma-t_1}{t_l-t_l}\tag{8}$$

where σ is a sufficiently small positive number, $t_1\leqslant\cdots t_i\cdots\leqslant t_l$ are the time when data samples are collected. In this model, the last data sample is thought to be the most important and its membership equals 1, the first data sample be the least important and its membership equals σ. Eq. (8) is called time-based model. This model lacks solid theoretical foundation and only suits for sequential learning problem. Time-based membership model has another form as follows[11]:

$$\mu_i=\frac{1}{1+\exp\left[a-2a\left(\dfrac{t_i-t_1}{t_l-t_1}\right)\right]}\tag{9}$$

Another type of membership model is space-based model which determines the fuzzy membership according to the spatial distribution of data samples. Huang proposed a membership model based on outlier detection method[10]

$$\mu_i=\begin{cases}1-\dfrac{\|\boldsymbol{x}_i-\overline{\boldsymbol{x}}\|}{\max\limits_j(\|\boldsymbol{x}_j-\overline{\boldsymbol{x}}\|)}+\sigma, & \boldsymbol{x}_i\in\boldsymbol{M}\\ v, & 0<v<\sigma,\quad \boldsymbol{x}_i=\boldsymbol{O}\end{cases}\tag{10}$$

The outlier set $\boldsymbol{O}$ is detected from training set, and the rest part of the training set is called main body set $\boldsymbol{M}$. $\overline{\boldsymbol{x}}$ is the cluster center of all data points in $\boldsymbol{M}$. For $\boldsymbol{x}_i\in\boldsymbol{M}$, the membership values are defined according to their distance to the center of the main body. For an outlier, the membership of sample is equal to a sufficiently small positive number v. In this model, outliers are equally treated and membership values of outliers are assigned the same values.

In this paper, strategies based on kNN method and SVDD are used to set the fuzzy membership values of data points.

3.1 Strategy of using kNN

Given two samples x_1, $x_2\in\boldsymbol{R}^n$, the distance between two samples in the feature space $\boldsymbol{H}$ is defined by

$$\begin{aligned}d(x_1,x_2)&=\|\varphi(x_1)-\varphi(x_2)\|_2\\&=\sqrt{K(x_1,x_1)-2K(x_1,x_2)+K(x_2,x_2)}\end{aligned}\tag{11}$$

For each point $\boldsymbol{x}_i$, we can find a set $\boldsymbol{S}_i^k$ that consists of k nearest neighbors of $\boldsymbol{x}_i$ according to the definition of Eq. (11). The average distance between $\boldsymbol{x}_i$ and each element of $\boldsymbol{S}_i^k$ is defined as

$$d_i=\frac{1}{k}\sum\sqrt{K(\boldsymbol{x}_i,\boldsymbol{x}_i)-2K(\boldsymbol{x}_i,\boldsymbol{x}_j)+K(\boldsymbol{x}_j,\boldsymbol{x}_j)}\tag{12}$$

For data points x_1 and x_6 in Fig. 2, S_1^4 and S_6^4 can be obtained as $S_1^4=\{x_2,x_3,x_4,x_5\}$ and $S_6^4=\{x_7,x_8,x_9,x_{10}\}$ according to Eq. (11). The average distance d_1 and d_6 can be calculated according to Eq. (12). From Fig. 2, we can see that d_1 is smaller than d_6.

We observe this situation and assume that the data point with a larger value of d_i can be considered as outlier and should make less contribution to the regression accuracy. For this assumption, we can build a relationship between the fuzzy membership μ_i and the value of d_i. The maximal and minimal value of d_i are defined as

$$d_{\max}=\max(d_i|\boldsymbol{x}_i\in\boldsymbol{\chi}),\ d_{\min}=\min(d_i|\boldsymbol{x}_i\in\boldsymbol{\chi})\tag{13}$$

The fuzzy membership function is defined as follows:

$$\mu_i = 1 - (1-\sigma)\left(\frac{d_i - d_{\min}}{d_{\max} - d_{\min}}\right)^f \tag{14}$$

where $\sigma < 1$. f is the parameter that controls the degree of mapping function as shown in Fig. 3.

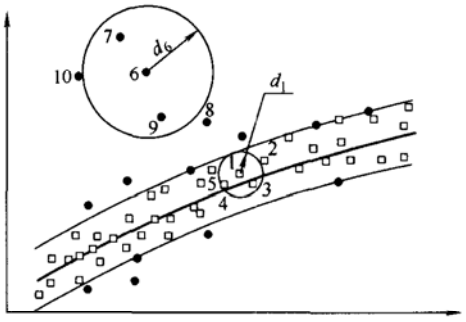


Figure 2 The values of d_1 and d_6 in SVR
● support vector; □ non-support vector;
—— regression line; ····· regression margin line

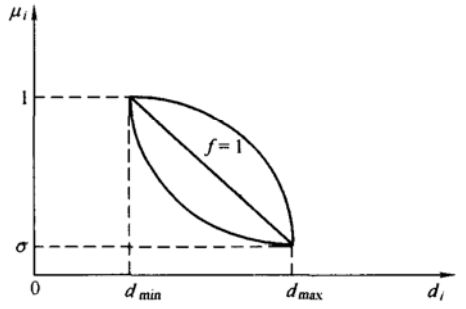


Figure 3 The fuzzy membership function based on *k*NN

When d_i is close to $d_{\min}$, data point $\boldsymbol{x}_i$ is near the regression line whose fuzzy membership μ_i is close to 1. With the increase of d_i, $\boldsymbol{x}_i$ is away from the regression line and the fuzzy membership of $\boldsymbol{x}_i$ will decrease. When d_i is close to $d_{\max}$, $\boldsymbol{x}_i$ can be regarded as an outlier and μ_i is close to a sufficiently small positive number σ. This can effectively reduce the effect of outliers.

3.2 Strategy of using SVDD

The idea of SVDD was introduced in Refs. [12,13]. Data samples in input space are mapped into a high dimensional feature space, then fuzzy membership value of each input sample is confirmed according to its distance to the center of the smallest enclosing hypersphere.

Let $T = \{x_1, \cdots, x_i, \cdots, x_l\}$ be a data set of l points, $\boldsymbol{x}_i \in \boldsymbol{R}^n$, $i = 1, \cdots, l$, in order to obtain the data domain description of training set, we try to find the smallest enclosing hypersphere which are described by center $\boldsymbol{a}$ and radius R. If the data samples are not spherically distributed in input space, we can map them into a high dimensional feature space *via* mapping Φ: $\boldsymbol{R}^n \to \boldsymbol{F}$, where $\boldsymbol{F}$ is the feature space. Then in the high dimensional feature space, we look for the smallest enclosing hypersphere by solving the QP problem as follows:

$$\min W(\xi_i, R, \boldsymbol{a}) = R^2 + C\sum_{i=1}^{l}\xi_i$$
$$\text{s.t.} \quad \begin{cases} \|\Phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2 \leqslant R^2 + \xi_i \\ \xi_i \geqslant 0, i, = 1, 2, \cdots, l \end{cases} \tag{15}$$

where C is slack factor which determines the trade off between the volume of the sphere and the number of target objects rejected. Introducing Lagrange multipliers β_i, β_j finally we obtain the Wolf dual as follows:

$$\max Q(\beta) = \sum_{i=1}^{l} K(\boldsymbol{x}_i, \boldsymbol{x}_i)\beta_i - \sum_{i,j=1}^{l}\beta_i\beta_j(\boldsymbol{x}_i, \boldsymbol{x}_j)$$
$$\text{s.t.} \quad 0 \leqslant \beta_i \leqslant C, i = 1, 2, \cdots, l \tag{16}$$

After solving the QP problem, we obtain the optimal Lagrange multipliers of each data sample. The distance between the data sample and the center of feature space hypersphere is defined as[12]

$$D^2(\boldsymbol{x}_i) = \sum_{i,j=1}^{l}\beta_i\beta_j K(\boldsymbol{x}_i, \boldsymbol{x}_j) + K(\boldsymbol{x}_i, \boldsymbol{x}_i) - 2\sum_{j=1}^{l} K(\boldsymbol{x}_j, \boldsymbol{x}_i)\beta_j \tag{17}$$

Then the radius of the smallest enclosing hypersphere is determined by $R = D(\boldsymbol{x}_i)|\forall 0 < \beta_i < C$. and the membership function is defined as

$$\mu_i = \begin{cases} \left(1 - \dfrac{D(\boldsymbol{x}_i) - D_{\min}}{D_{\max} - D_{\min}}\right)^f + \sigma, R < D(\boldsymbol{x}_i) \leqslant D_{\max} \\ 1 - \dfrac{D(\boldsymbol{x}_i) - D_{\min}}{D_{\max} - D_{\min}}, \quad D_{\min} \leqslant D(\boldsymbol{x}_i) \leqslant R \end{cases} \tag{18}$$

where $D_{\max}$, $D_{\min}$ are maximal and minimal values of $D(\boldsymbol{x}_i)$, repectively. The membership function is shown in Fig. 4.

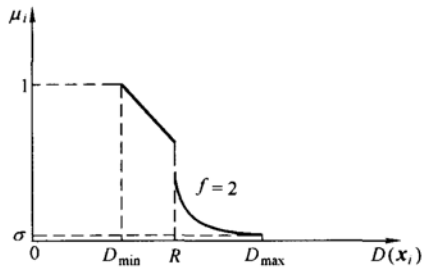


Figure 4 Fuzzy membership changing with distance form the center of the smallest enclosing sphere

For a point $\boldsymbol{x}_i$ in input space, distance between mapping $\Phi(\boldsymbol{x}_i)$ and the center of the small enclosing hypersphere satisfies $D_{\min} \leqslant D(\boldsymbol{x}_i) \leqslant D_{\max}$, when $D_{\min} \leqslant D(\boldsymbol{x}_i) \leqslant R$, $\boldsymbol{x}_i$ belongs to the data description. The data point is near the regression margin, such as the points 3 and 4 in Fig. 2. With the increase of $D(\boldsymbol{x}_i)$, their membership values will linearly decrease. When $R < D(\boldsymbol{x}_i) \leqslant D_{\max}$, point $\boldsymbol{x}_i$ is far away from the regression margin, shown as the points 6, 7 in Fig. 2. Membership value of $\boldsymbol{x}_i$ is a quadratic function of $D(\boldsymbol{x}_i)(f = 2)$. With the increase of $D(\boldsymbol{x}_i)$, membership value of $\boldsymbol{x}_i$ rapidly increases. When $D(\boldsymbol{x}_i)$ is closing to $D_{\max}$, μ_i is closing to a sufficiently small positive number σ.

From the definition of kNN-based FSVR and SVDD-based FSVR, we can see that both of them need to compute the kernel functions during the process of determining fuzzy membership model. But an extra QP problem needs to be solved in SVDD-based FSVR [see Eq. (16)]. Thus the SVDD-based FSVR is more time-consuming than kNN-based FSVR. As discussed in Ref. [8], the QP problem of SVM spends most of the training time on calculating kernel functions. Hence if we cache the most usually used kernel functions into memory during the process of determining fuzzy membership model, the training time of SVDD-based FSVR will be significantly reduced as it does not need to calculate kernel functions again in the process of training.

4 SIMULATION RESULTS

According to the prior knowledge and experience, the key variables affect the oxidation process as well as the concentration of 4-CBA, are the residence time of reactants, the ratio of PX to acetic acid, the ingredient and concentration of catalyst, reaction temperature and pressure, the partial pressure of oxygen and water content in the reactor. After the comparison between those key variables and industrial process variables, 10 process variables are selected as input variables of the soft sensor model[2], as listed in Table 2. The schematic layout of PX oxidation process in Fig. 1 shows that there exists a substantial time delay between each process variable and the product quality variable. Every tank has a residence time varying from 15 min to 71 min. The total time delay of the process is about 200 min.

A total of 403 data samples were collected including normal operation and load down operation from the distributed control system (DCS) of a practical PX oxidation process. These data were divided into two sets, with one set having 171 samples to be used for training, and the other having 232 samples for testing. Firstly, we trained the SVR in the training set, with kernel function being Gauss RBF kernel with parameters $p = 8$ and $C = 64$ obtained by cross-validation[5]. The predicted result on the testing set is shown in Fig. 5(a). Root-mean-square-error (RMSE) is shown in Table 3.

Table 2 The all variables of process model

No.	Variables	Dead time min	Normal value
input variables			
1	para-xylene to feed mixing tank, kg·h^{-1}	205	28243.57
2	feed to reactor, kg·h^{-1}	180	143008.71
3	catalyst concentration, μg·g^{-1}	185	361.07
4	reactor temperature, ℃	110	188.05
5	level of reactor, %	110	88.99
6	reactor condenser to water withdraw, kg·h^{-1}	95	673.23
7	vent O_2 concentration from the reactor, % (by volume)	95	4.08
8	total water withdrawal, kg·h^{-1}	90	40939.27
9	first crystallizer temperature, ℃	75	184.88
10	vent O_2 concentration from the first crystallizer, % (by volume)	70	4.07
output variable			
11	the concentration of 4-CBA in the crude TA, μg·g^{-1}	0	2702

In order to obtain the parameters of kNN-based FSVR, for a point $\boldsymbol{x}_i$, we employed kNN algorithm to calculate d_i to obtain $d_{\min} = 8.6482$ and $d_{\max} = 9.3754(k = 32)$. Let $f = 2$ and $\sigma = 0.01$, according to Eq. (14), we obtained the membership function as follows:

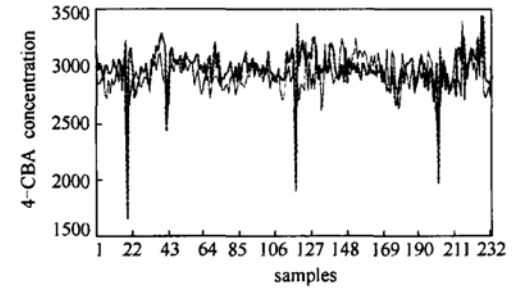
$$\mu_i = -1.87d_i^2 + 32.38d_i - 139.02 \tag{19}$$

To find the parameters of SVDD-based FSVR, we used SVDD algorithm to train on the data set. Kernel function is Gauss RBF kernel with $p = 10$ and $C = 0.1$. After training we obtained $R = 12.648$ and $D_{\max} = 12.661$ and $D_{\min} = 12.407$. Let $f = 2$ and $\sigma = 0.01$, according to Eq. (18), we obtained the membership model as follows:

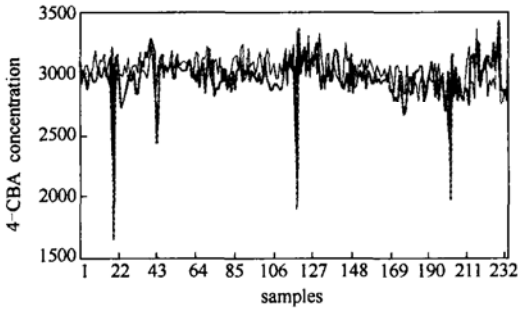
$$\mu_i = \begin{cases} 15.50D^2(\boldsymbol{x}_i) - 392.49D(\boldsymbol{x}_i) + 2484.68, \\ \qquad 12.648 < D(\boldsymbol{x}_i) \leqslant 12.661 \\ -3.94D(\boldsymbol{x}_i) + 49.85, \\ \qquad 12.407 < D(\boldsymbol{x}_i) \leqslant 12.648 \end{cases} \tag{20}$$

Using the obtained membership models, we started to train kNN-based and SVDD-based FSVR on the training set, respectively. Kernel function and slack

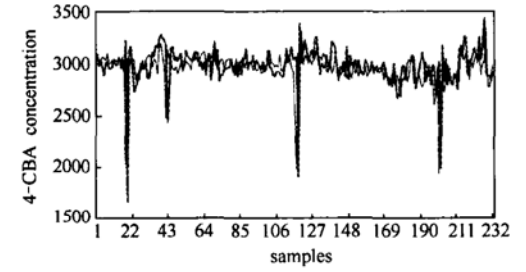
factor are the same as in SVR. The predicted results are given in Figs. 5(c) and (d).



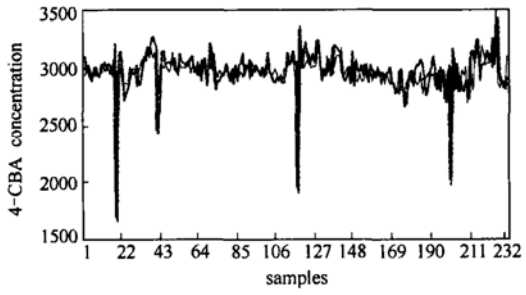
(a) SVR



(b) Time-based FSVR



(c) kNN-based FSVR



(d) SVDD-based FSVR

Figure 5 Predicted results of SVR, time-based FSVR, kNN-based FSVR and SVDD-based FSVR
····· real values; —— predicted results

We can see that the accuracy of kNN-based or SVDD-based FSVR is higher than that of SVR. Because of the influence of outliers which are produced by load down operation in PX oxidation process, SVR does not work well and its accuracy is expected to be lower. On the contrary, in kNN-based and SVDD-based FSVR, different data samples are assigned with different fuzzy membership values. Thus the predicted results of kNN-based and SVDD-based FSVR fit the real value curve very well, and their accuracies are higher than that of SVR.

Table 3 RMSE of different algorithms

Algorithm	RMSE
SVR	0.1071
time-based FSVR	0.08911
kNN-based FSVR	0.06292
SVDD-based FSVR	0.05173

Finally, we trained the time-based FSVR on the training set with the same kernel parameters and slack factor as in kNN-based FSVR. Fig. 5(b) presents the predicted result. From Table 3 we can see that the accuracy of time-based FSVR is lower than that of kNN-based and SVDD-based FSVR. It indicates that time-based membership model is not suitable for the case when the training set includes noises.

5 CONCLUSIONS

In this paper, a fuzzy support vector regression method is presented to deal with over-fitting problem caused by outliers. Membership models based on kNN and SVDD method are also proposed to fuzzify all the training data. The proposed soft sensor models based on kNN and SVDD are applied to predict the concentration of 4-CBA in a PTA production process. Simulation results indicate that the proposed methods indeed reduce the effect of outliers and yield higher accuracy than either SVR or time-based FSVR.

NOMENCLATURE

C	slack variables
d	coefficient of polynomial kernel
$d(\boldsymbol{x}_1, \boldsymbol{x}_2)$	distance between the samples $\boldsymbol{x}_1, \boldsymbol{x}_2$
f	degree of mapping function
$\boldsymbol{H}$	feature space
$K(g, g)$	kernel function
L_j	class label of training sample
l	number of training samples
$\boldsymbol{M}$	main body set
$\boldsymbol{O}$	outlier set
p	scale parameter of Gaussian kernel
$\boldsymbol{R}$	input space
t_i	time when data samples are collected
$\boldsymbol{w}$	weight value vector
$\boldsymbol{x}_i$	training sample
$\overline{\boldsymbol{x}}$	cluster center of all data points
y_i	class label of training sample
α_i, α_i^*	Lagrange coefficients
ξ_i, ξ_i^*	slack variables
μ_i	fuzzy membership values of sample $\boldsymbol{x}_i$
σ	sufficiently small positive number
$\boldsymbol{\chi}$	sample set in input space

REFERENCES

1 Mu, S.J., Su, H.Y., Gu, Y., Chu, J., "Multi-objective optimization of industrial PTA oxidation process", *Chinese J. Chem. Eng.*, **11** (5), 536—541 (2003).

2 Wang, L.J., "Studies on the kinetics of *p*-xylene oxidation and the reactor simulation", Dissertation for Master Degree, Zhejiang Univ., Hangzhou (2001). (in Chinese)

3 Cao, G., Servida, A., Pisu, M., Morbidelli, M., "Kinetics of *p*-xylene liquid-phase catalytic oxidation", *AIChE J.*, **40**, 1156—1166(1994).

4 Liu, R.L., Su, H.Y., Mu, S.J., Jia, T., Cheng, W.Q., Chu, J., "Fuzzy neural network model of 4-CBA concentration for industrial PTA oxidation process", *Chinese J. Chem. Eng.*, **12** (2), 234—239 (2004).

5 Vapnik, V., The Nature of Statistical Learning Theory, Springer, New York(1995).

6 Guyon, I.,Weston, J., Barnhill, S., Vapnik, V., "Gene selection for cancer classification using support vector machines", *Machine Learning*, **46** (6), 389—422 (2002).

7 Cortes, C., Vapnik, V., "Support vector networks", *Machine Learning*, **20**, 273—297 (1995).

8 Burges, C.J.C., "A tutorial on support vector machines for pattern recognition", *Data Mining and Knowledge Discovery*, **2** (2), 1—47 (1998).

9 Lin, C.F., Wang, S.D., "Fuzzy support vector machines", *IEEE Trans. on Neural Networks*, **13** (3), 466—471 (2002).

10 Huang, H.P., Liu, Y.H., "Fuzzy support vector machines for pattern recognition and data mining", *International J. Fuzzy Systems*, **14** (3), 826—834 (2002).

11 Zhang, X., "Using class center vectors to build support vector machines", In: Proceedings of NNSP'99, Madison, USA (1999).

12 Tax, D.M.J., Duin, R.P.W., "Data domain description by support vectors", In: Proceedings of ESANN'99, Brussels, Belgium (1999).

13 Ben-Hur, A., Horn, D., Siegelmann, H.T., Vapnik, V., "Support vector clustering", *J. Machine Learning Research*, **2**, 125—137 (2001).