Research Article

Fault Isolation for Nonlinear Systems Using Flexible Support Vector Regression

Yufang Liu,[1,2] Bin Jiang,[1] Hui Yi,[3] and Cuimei Bo[3]

[1] *College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, No. 29 Yudao Street, Nanjing 210016, China*
[2] *Department of Technology Research, Guodian Environment Protection Research Institute, No. 10 Pudong Street,*
  Nanjing 210032, China
[3] *College of Automation and Electronics, Nanjing University of Technology, No. 30 Puzhu South Road, Nanjing 211816, China*

Correspondence should be addressed to Bin Jiang; binjiang@nuaa.edu.cn

While support vector regression is widely used as both a function approximating tool and a residual generator for nonlinear system fault isolation, a drawback for this method is the freedom in selecting model parameters. Moreover, for samples with discordant distributing complexities, the selection of reasonable parameters is even impossible. To alleviate this problem we introduce the method of flexible support vector regression (F-SVR), which is especially suited for modelling complicated sample distributions, as it is free from parameters selection. Reasonable parameters for F-SVR are automatically generated given a sample distribution. Lastly, we apply this method in the analysis of the fault isolation of high frequency power supplies, where satisfactory results have been obtained.

1. Introduction

With the increasing use of complex systems, there has been great interest in the development of techniques to fault isolations. Generally, the major approaches for fault isolation can be divided into two categories, namely, model-based and data-driven techniques. The fundamental aspect of a model-based fault isolation is a process model that runs parallel to the process [1]. With traditional methods like observers, approximating the function between state vectors and input/output vectors is successful due to precise mathematical modelling by the use of filters. While these methods have successfully modelled linear systems, when applied to nonlinear systems like chemical processing, precise devices, and aerodynamic systems, they often fail to construct a sufficient model because their mechanism models are hard to be formed. Model-based approaches have advantages in terms of on-board implementation considerations, but their reliability may decrease as the nonlinear system complexities increase [2].

Therefore, data-driven techniques have been introduced to more accurately construct process models as these methods are free from the requirement to analytically derive equations for a given system, shown in Figure 1. One feasible method is to use the artificial neural network (ANN). ANN utilizes experience risk minimization (ERM) principle to construct the process model, where the target function is numerically approximated by minimizing residuals between function estimates and outputs of the process data. Applications of ANN-based fault isolation have been widely addressed in the literature. For example, Sadough Vanini et al. [2] used the dynamic neural networks to isolate the fault of a dual spool gas turbine engine. Filippetti et al. [3] applied fuzzy-NN to the fault isolation of induction motor drives. However, the learning ability of ANN is dependent on the number of training samples. It requires massive samples to ensure the modelling performance. But in most practical applications, not many of fault samples can be acquired.

More recently, the principle of structure risk minimization [4] has been introduced in fault isolation through the utilization of support vector regression (SVR) [5, 6] as it can provide more accurate results than using neural networks in condition of smaller training samples. It was constructed on the basis of statistics learning theory that provides

the theoretical proofs of learning from finite samples. Much has been addressed in the literature where SVR shows superiorities to ANN in process modeling [7].

However, the performance of SVR-based modelling is greatly affected by its parameters. Although SVR has been well studied and many remarkable achievements have been obtained, the theoretical estimation of regression parameter remains unsolved in the last decade. There is no general consensus on the selection of proper parameters, but only some practical recommendations on this issue. This greatly increases the difficulty for common operators to master the SVR-based approach. Moreover, in some complicated cases, there are even no reasonable parameter settings that could be found. A rigorous selection of regression parameters can lead to the overlearning of training samples, while slack selections can lead to underlearning. There exist no parameters that yield good trade-off between overlearning and underlearning.

In this paper we introduce a flexible SVR (F-SVR) approach [8] to more accurately implement models that construct different residual generators for fault isolation. By automatically dividing training samples into several regions, this method is not only free from parameter selection, but also able to learn well and to generalize well for complicated cases.

2. Problem Statement

Support vector regression (SVR) is a process modeling tool that approximates the function between inputs and outputs:

$$y = f(x) = wx + b. \tag{1}$$

Here x and y represent the input and output vectors, respectively, f is the modeled function, w is its weight vector, and b is the bias decided by the vector w.

The SVR-based modelling can be viewed as a process of finding the optimal weight vector w_0 with a proper parameter vector α_0 for a given data set $\{(x_1, y_1), \ldots, (x_n, y_n)\}$:

$$\begin{aligned}
(\alpha_0, w_0) &= \arg\min_{\alpha, w}: R_{SRM}(\alpha, w) \\
&= \arg\min_{\alpha, w}: \text{Residual} + \phi(w) \\
&= \arg\min_{\alpha, w}: \int_{T_o}^{T_k} L(y, f(x, \alpha, w)) \, p(x, y) \, dx \, dy,
\end{aligned} \tag{2}$$

where $L(y_i, f(x_i, w, \alpha)) = C_i \cdot |y_i - (\sum_{i=1}^{l} \beta_i K(x, x_i) + b)|_\varepsilon$ is the loss function, $p(x, y)$ is the unknown joint distribution of x and y, C_i is the regularized parameter, ε is the insensitive parameter, $K(\cdot)$ is the kernel function, $\phi(w)$ denotes the generalization ability for the regression, and $\alpha = \{C, \varepsilon, K(\cdot)\}$. The optimal weight vector w_0 could be obtained by Lagrangian approaches. Thus the core problem for the SVR modeling is the selection of parameters.

The framework for SVR-based fault isolation is shown in Figure 2. Different operating models are constructed by
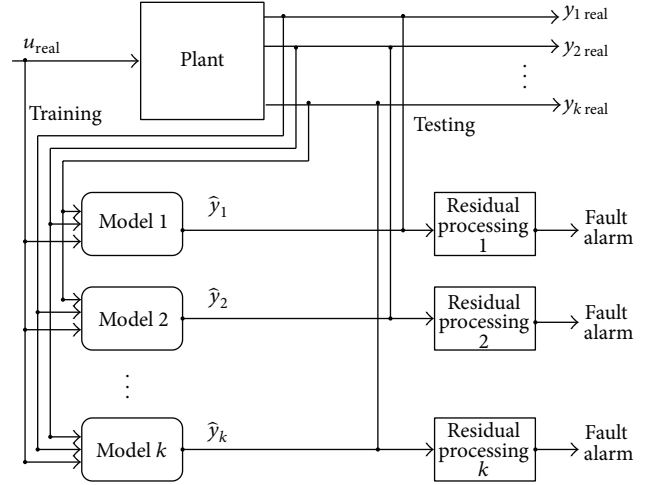

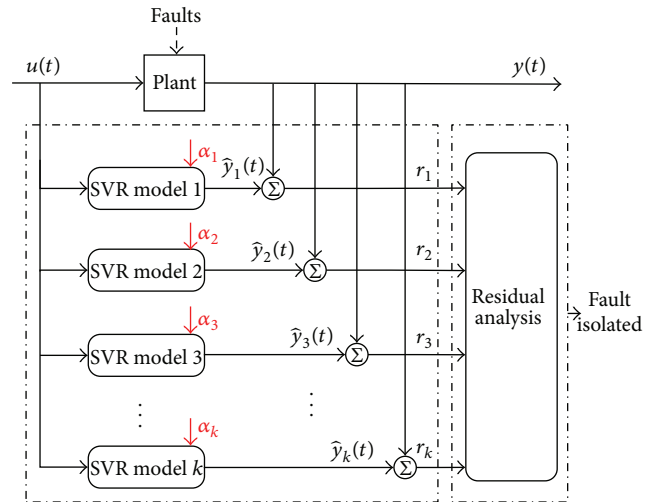Figure 1: Framework of data-driven fault isolation.

Figure 2: Scheme of SVR-based fault isolation.

SVRs with given parameters. The residuals between estimated outputs and real outputs are generated for fault isolation.

However, the selection of SVR parameters is not easy. With the fixed regression parameter α, the drawback of SVR is the hardness of the trade-off between overlearning and underlearning. Moreover, in some complicated cases, even no reasonable parameters could be found. As shown in Figure 2, due to the different complexities of sample distributions, the requirements of parameters are discordant. If a rigorous parameter α_1 is selected, the regression is overlearning in region A. In contrast, if a slack parameter α_2 is selected, the regression fails to learn in region B (underlearning). No parameter that can adequately fit all of the cases exists. Consequently, we advocate using the F-SVR approach in order to overcome this drawback [8], as the regression parameter is automatically generated and is variable instead of fixed (Figure 3).
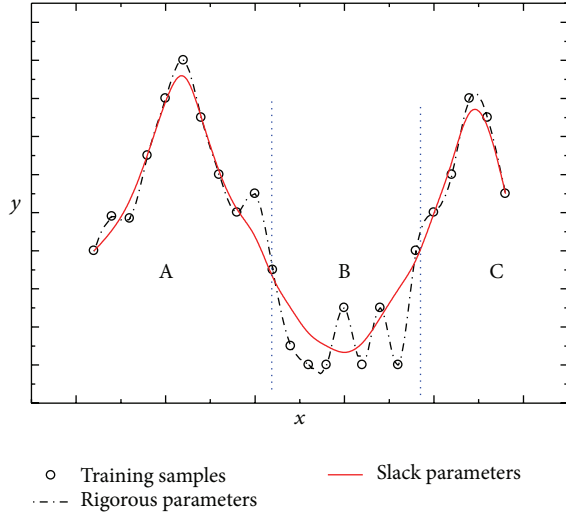
FIGURE 3: The drawback for SVR modeling.

○ Training samples ── Slack parameters
·-·- Rigorous parameters

3. Fault Isolation Using Flexible Support Vector Regression (F-SVR)

3.1. The Principle of Flexible Support Vector Regression. We proposed a parameter-free algorithm for process modeling, namely, flexible support vector regression. The F-SVR attempts to divide the training samples into k regions according to the distribution complexity, and for the ith region, parameter α_i is generated. By minimization (3), the function between x and y is approximated:

$$
\begin{aligned}
R(w, \alpha) = &\iint_{T1} L\left(y, f\left(x, w, \alpha_1\right)\right) p(x, y) \, dx \, dy \\
&+ \iint_{T2} L\left(y, f\left(x, w, \alpha_2\right)\right) p(x, y) \, dx \, dy \\
&+ \cdots + \iint_{Tk} L\left(y, f\left(x, w, \alpha_k\right)\right) p(x, y) \, dx \, dy.
\end{aligned}
\tag{3}
$$

With given parameters, (3) can be minimized by solving a *quadratic programming* (QP) problem. Supposing the training samples were divided into two areas, $(x_i, y_i)_{i=1}^{q}$ and $(x_i, y_i)_{i=q+1}^{l}$, and $\alpha = \{C, \sigma, \varepsilon\}$ are parameters for the two areas, the minimization of (3) could be termed as

$$
\begin{aligned}
\text{min:} \quad & \frac{1}{2}(w, w) + C_1\left(\sum_{i=1}^{q} \xi_i^* + \sum_{i=1}^{q} \xi_i\right) \\
& + C_2\left(\sum_{i=q+1}^{l} \xi_i^{*2} + \sum_{i=q+1}^{l} \xi_i^2\right)
\end{aligned}
\tag{4}
$$

$$
\begin{aligned}
\text{S.t.} \quad & y_i - (w, x_i) - b \leq \varepsilon + \xi_i^* \\
& (w, x_i) + b - y_i \leq \varepsilon + \xi_i \\
& \xi_i^* \geq 0; \quad \xi_i \geq 0.
\end{aligned}
$$

And the Lagrangian for (4) is

$$
\begin{aligned}
&L\left(w, \xi^*, \xi, \alpha^*, \alpha, \gamma^*, \gamma\right) \\
&= \frac{1}{2}(w, w) + C_1\left(\sum_{i=1}^{q} \xi_i^* + \sum_{i=1}^{q} \xi_i\right) \\
&\quad + C_2\left(\sum_{i=q+1}^{l} \xi_i^{*2} + \sum_{i=q+1}^{l} \xi_i^2\right) \\
&\quad - \sum_{i=1}^{l} \alpha_i\left[y_i - w \cdot x_i - b + \varepsilon + \xi_i\right] \\
&\quad - \sum_{i=1}^{l} \alpha_i^*\left[w \cdot x_i + b - y_i + \varepsilon + \xi_i^*\right] \\
&\quad - \sum_{i=1}^{l}\left(\gamma_i^* \xi_i^* + \gamma_i \xi_i\right).
\end{aligned}
\tag{5}
$$

Taking the partial derivative for (5), we get

$$
\begin{aligned}
\frac{\partial L}{\partial w} &= w + \sum_{i=1}^{l} \alpha_i x_i - \sum_{i=1}^{l} \alpha_i^* x_i = 0, \\
\frac{\partial L}{\partial b} &= \sum_{i=1}^{l}\left(\alpha_i - \alpha_i^*\right) = 0, \\
\frac{\partial L}{\partial \xi_i} &= \begin{cases} C_1 - \alpha_i - \gamma_i, & 1 \leq i \leq q \\ 2C_2 \xi_i - \alpha_i - \gamma_i, & q < i \leq l \end{cases} = 0, \\
\frac{\partial L}{\partial \xi_i^*} &= \begin{cases} C_1 - \alpha_i^* - \gamma_i^*, & 1 \leq i \leq q \\ 2C_2 \xi_i^* - \alpha_i^* - \gamma_i^*, & q < i \leq l \end{cases} = 0.
\end{aligned}
\tag{6}
$$

Make the dual problem for (4):

$$
\begin{aligned}
\text{max:} \quad & -\varepsilon \sum_{i=1}^{l}\left(\alpha_i^* + \alpha_i\right) + \sum_{i=1}^{l} y_i\left(\alpha_i^* - \alpha_i\right) \\
& - \frac{1}{2} \sum_{i,j=1}^{l}\left(\alpha_i^* - \alpha_i\right)\left(\alpha_j^* - \alpha_j\right)\left(x_i \cdot x_j\right) \\
& - \sum_{i=q+1}^{l} \frac{1}{4C_2}\left(\alpha_i^2 + \alpha_i^{*2}\right)
\end{aligned}
\tag{7}
$$

$$
\text{S.t.} \quad \sum_{i=1}^{l}\left(\alpha_i^* - \alpha_i\right) = 0; \quad 0 \leq \alpha_i^* \leq C_i; \quad 0 \leq \alpha_i \leq C_i,
$$

where

$$
C_i = \begin{cases} C_1, & 1 \leq i \leq q \\ C_2, & q+1 \leq i \leq l. \end{cases}
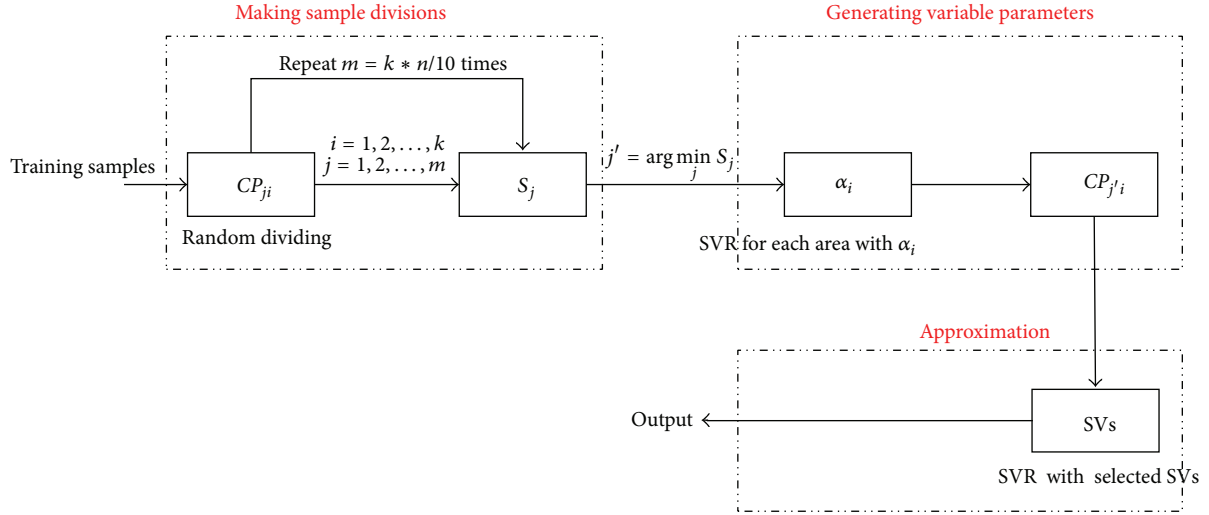\tag{8}
$$

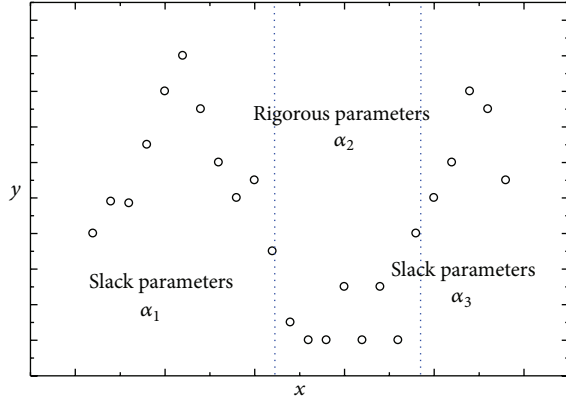FIGURE 4: The flowchart for F-SVR modeling.

FIGURE 5: Setting different parameters for different regions.

As $\alpha \cdot \alpha^* \equiv 0$, (7) could be written with the following form:

$$\text{min:} \quad \frac{1}{2}(\alpha - \alpha^*)^T Q (\alpha - \alpha^*) + y^T (\alpha - \alpha^*) \\ + \varepsilon e^T (\alpha + \alpha^*) - (\alpha - \alpha^*)^T v (\alpha - \alpha^*), \tag{9}$$

where $Q = K(x_i, x_j)$, $v = [0, \ldots, (1/4)C_{q+1}^{-1}, \ldots, (1/4)C_l^{-1}]$, and e is the unit vector. Further, the regression could be written as the QP problem

$$\text{min:} \quad \frac{1}{2}\left[\alpha^T, (\alpha^*)^T\right]\begin{bmatrix} D & -D \\ -D & D \end{bmatrix}\begin{bmatrix} \alpha \\ \alpha^* \end{bmatrix} + c^T \begin{bmatrix} \alpha \\ \alpha^* \end{bmatrix} \tag{10}$$

$$\text{S.t.} \quad z^T \begin{bmatrix} \alpha \\ \alpha^* \end{bmatrix} = 0; \quad 0 \le \alpha_i \le C_i; \quad 0 \le \alpha_i^* \le C_i,$$

where $D = Q - v$, $c = [\varepsilon e + y, \varepsilon e - y]$, and

$$z_i = \begin{cases} 1, & i = 1, 2 \ldots q \\ -1, & i = q+1, \ldots, l. \end{cases} \tag{11}$$

This form of QP problem can be solved by an active set method [9]. The feasibility for complicated cases that more regions are divided can similarly be proved.

3.2. Detailed Process of F-SVR Modeling. F-SVR modeling contains three major steps, shown in Figure 4.

Step 1 (sample divisions). This section shows how T_i in (3) is determined. Given training samples $(x_i, y_i)_{i=1}^n$, formula (12) is utilized to estimate the distribution complexity

$$CP = \frac{\sum_{i=1}^{n-1} |A_i| / (n-1)}{\sum_{i=1}^{n-2} \cos \theta_i / (n-2)} = \frac{n-2}{n-1} \cdot \frac{\sum_{i=1}^{n-1} |A_i|}{\sum_{i=1}^{n-2} \cos \theta_i}, \tag{12}$$

where $A_i = (x_{i+1}, y_{i+1}) - (x_i, y_i)$ and $\cos \theta_i = |A_i \cdot A_{i+1}| / (|A_i| \cdot |A_{i+1}|)$. Supposing the training samples have been divided into k areas, formula (13) is implemented to evaluate the performance of division

$$S = \sum_{i=1}^{k} \frac{-C_i}{\sum_{j=1}^{k} C_j} \log \frac{C_i}{\sum_{j=1}^{k} C_j}. \tag{13}$$

The samples are divided randomly for several times, and the division with the smallest S value is treated as the best division:

$$j^* = \arg \min_j S_j, \quad j = 1, 2, \ldots m, \tag{14}$$

where m is number of times that random division is made and n is the number of training samples. And in this paper, we set $m = (k * n)/10$. The ith region denotes $x \in T_i$.

Step 2 (setting parameters for each region). Once the best division j^* is obtained, the CP values for all areas $CP_{j^*} = \{CP_{j^*1}, CP_{j^*2}, \ldots, CP_{j^*k}\}$ can also be obtained. In flexible

(a) The selection of support vectors
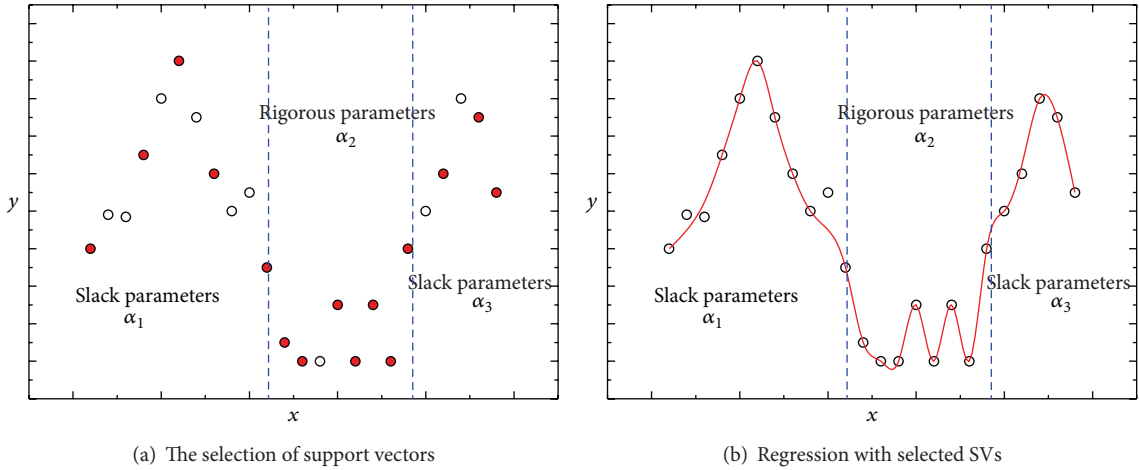

(b) Regression with selected SVs

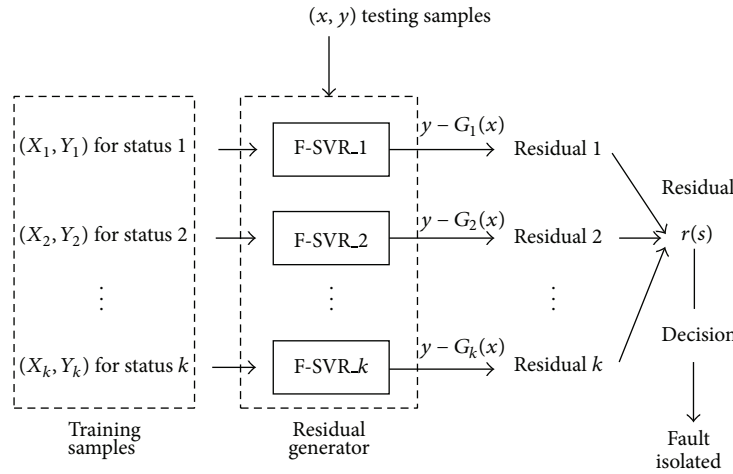FIGURE 6: F-SVR modeling for complicated distributions.

FIGURE 7: Scheme of F-SVR-based fault isolation.

support vector regression approach, the following empirical formulas are given to set the hyperparameter $\alpha_i = \{\varepsilon_i, C_i, \sigma_i\}$:

$$\alpha_i : \begin{cases} \varepsilon_i = \dfrac{0.5 * \text{std}^2 (X_i)}{CP_{j^*i}} \\ C_i = \dfrac{1000 CP_{j^*i}}{\sum_{i=1}^{k} CP_{j^*i}} \\ \sigma_i = \dfrac{5}{CP_{j^*i}}, \end{cases} \quad (15)$$

where $X_i = \{x \setminus x \in \text{the } i\text{th region}\}$.

Remark 1. The empirical setting of parameters for each region is referred to in Cherkassky's work [10] in 2004.

Step 3 (function approximation using selected support vectors). We use the conventional SVR with parameters α_i to extract informative samples for the ith region. As shown in Figure 6(a), the red samples are selected as support vectors (SVs). If m samples are selected as SVs for the training set $(x_i, y_i)_{i=1}^{n}$, the regression problem $f(x) = \sum_{i=1}^{n} \beta_i K(x, x_i) + b_0$

can be approximated by the regression problem of the SVs [8]:

$$f(x) = \sum_{i=1}^{m} \gamma_i K(x, \text{SV}_i) + b'_0. \quad (16)$$

Thus, the minimization of (3) can be simplified as

$$w^* = \arg \min_w : \int_{T_o}^{T_k} L(y, f(\text{SV}, \alpha, w)) p(\text{SV}, y) d_{\text{SV}} \cdot dy. \quad (17)$$

In (17), rigorous parameters $\alpha = \{0, 1000, 0.01\}$ are set to ensure the learning ability of regression. This problem can be solved by using the Lagrangian method in the same way that is used in conventional SVR (Figure 5).

As is shown in Figure 6(b), F-SVR (unlike conventional SVR) successfully approximates the distribution function without overlearning or underlearning. As we mentioned earlier, F-SVR is free from the issues that can arise in the manual selection of parameters as all of the parameters are generated automatically.
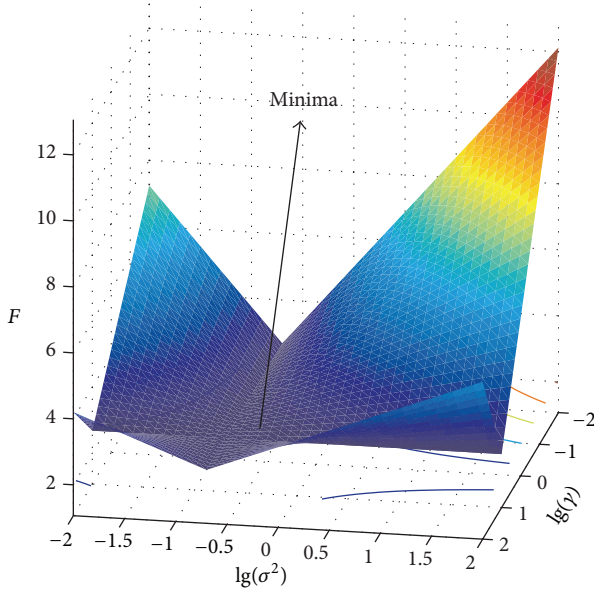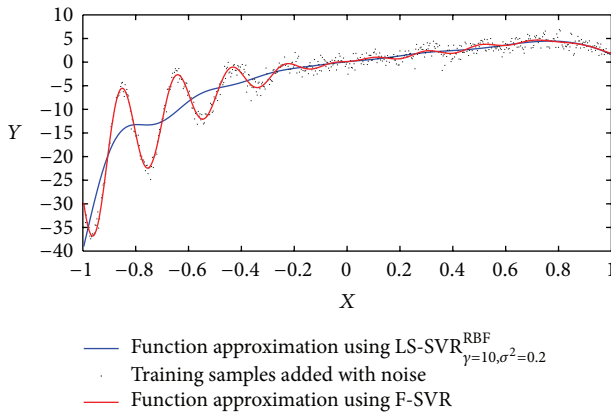
FIGURE 8: Parameter settings for LS-SVR.

—— Function approximation using LS-SVR$_{\gamma=10,\sigma^2=0.2}^{RBF}$
· Training samples added with noise
—— Function approximation using F-SVR

FIGURE 9: Comparison of learning ability.

3.3. The Basic Scheme of F-SVR-Based Fault Isolation. Fault isolation algorithms attempt to reveal which fault is occurring in the operating system. In our method, we determine which model of the system is most likely to be accurate by initially constructing the modes for all faulty statuses and then calculating the deviations between the real outputs and all of the model outputs. As is shown in Figure 7, once the best model has been located, the fault type can then be isolated.

Compared with conventional SVR-based fault isolation, the most significant contribution of our work is that we have alleviated the problem of parameter setting. What is required for F-SVR-based fault isolation are only process samples for different operating cases.

The detailed process of F-SVR-based fault isolation is as follows. Given $(x_i, y_i)_{i=1}^k$, $x_i, y_i \in R^n$ as the training samples from k different operating statuses concluding the normal

status and all faulty statuses, where (x_i, y_i) represents the training samples for the ith status, the function $G_i(\cdot)$ between input vector x_i and output vector y_i is initially approximated by the F-SVR method. After the training samples (x, y) construct the input/output models for all of the statuses, they are sent to these models to generate the residuals between the real outputs and the model outputs, thereby forming the residual vector:

$$\text{Residual}_i = y - G_i(x), \quad i = 1, \dots, k,$$

$$r = \begin{bmatrix} \text{Residual\_1} \\ \vdots \\ \text{Residual\_}k \end{bmatrix}_{k \times n}. \tag{18}$$

We define a function $R(i)$ to measure the deviation between the real output and the model output of the ith status

$$R(i) = \sum |y - G_i(x)|. \tag{19}$$

Faults can be isolated by analysis of the residual vector. In this paper, we simply regard the testing samples belonging to the i^* th status:

$$i^* = \arg \min_i R(i), \quad i = 1, \dots, k. \tag{20}$$

4. Experiments and Real Applications

4.1. Numerical Experiments. In our first attempt to validate our method by a numerical experiment, we used a data set with complicated distributions. White noise (SNR = 30 db) is added to the analytical equation shown in (21), where the training set consists of 600 samples extracted from $x \in (-1, 1]$. In order to approximate the distribution, both F-SVR and least-square SVR (LS-SVR) [11] are implemented:

$$y1 = (x - 0.5)^2 + 4 \sin(3x^2) + x,$$

$$y2 = 4x(1 - x)(2 \sin(30x + 24) + 3),$$

$$t = \frac{(1 - x)}{2}, \tag{21}$$

$$y = y1(1 - t) + y2 * t; \quad x \in (-1, 1].$$

To provide a fair comparison, the parameters for LS-SVR were optimized first using a grid search strategy. The evaluating index for LS-SVR is

$$F = \frac{d}{N} \cdot \sqrt{\frac{\sum_{i=1}^N (\widehat{y}_i - y_i)^2}{N - 1}}, \tag{22}$$

where d is the number of SVs and $\widehat{y}$ is the function output of LS-SVR. A small F value implies that the regression produces generalization and accurate learning. As is shown in Figure 8, the best parameter set for LS-SVR is $\{\gamma = 10, \sigma^2 = 0.2\}$.

The performance of the two methods is shown in Figure 9. It is clear in this case that F-SVR is a more accurate method for modeling data compared to the well-known LS-SVR.
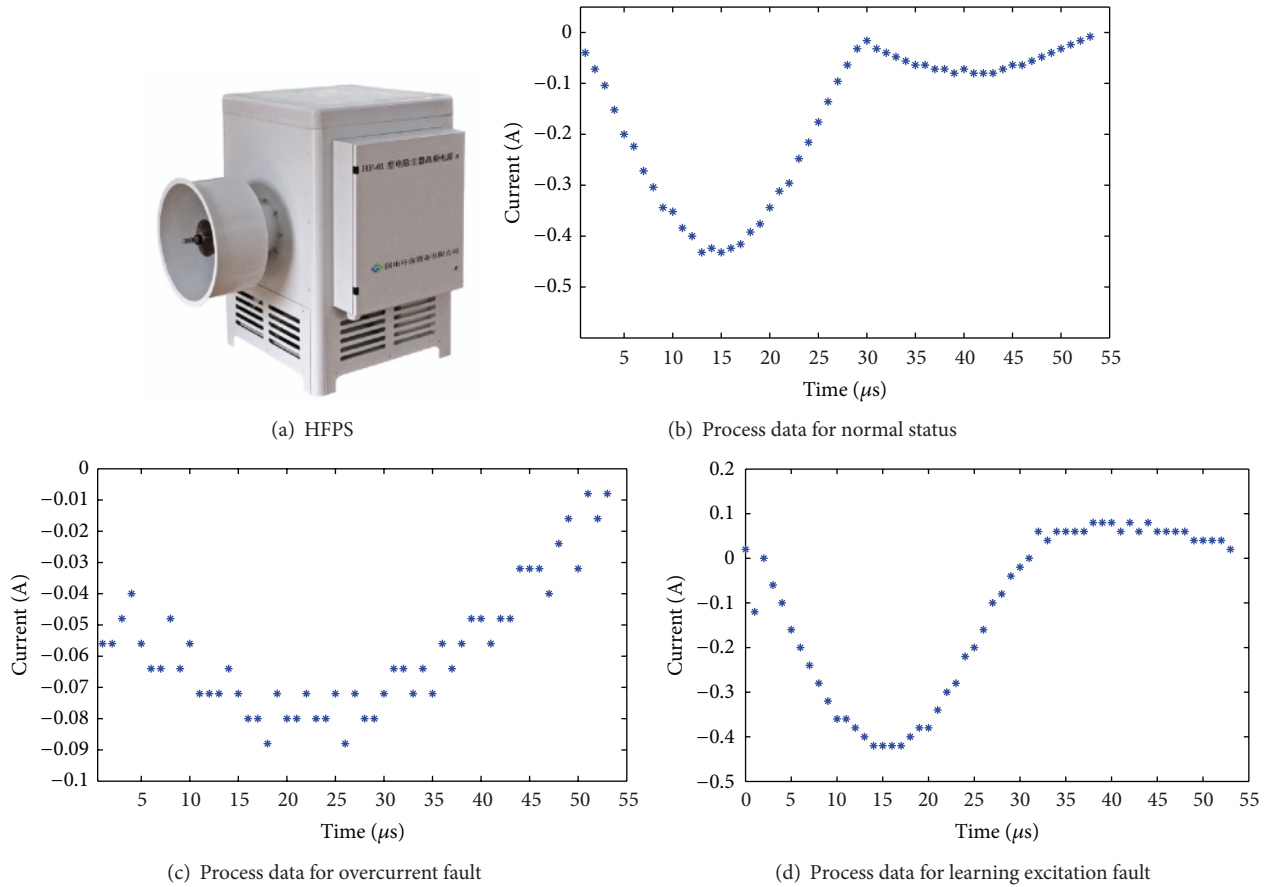
(a) HFPS

(b) Process data for normal status

(c) Process data for overcurrent fault

(d) Process data for learning excitation fault

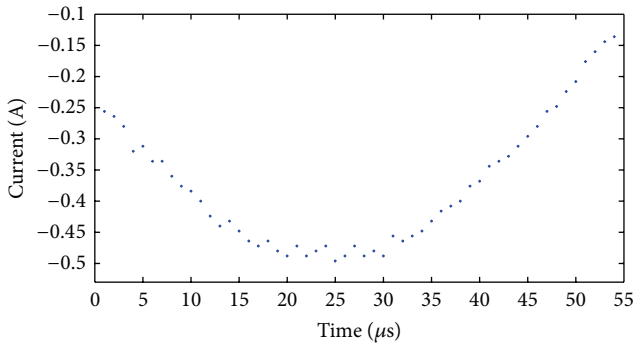Figure 10: Process data for different operating cases of HFPS.

Figure 11: The testing samples.

4.2. Fault Isolation for High Frequency Power Supply. The high frequency power supply (HFPS) is a nonlinear device that has been widely used in power plants for dedusting purpose. As its structure is highly complicated, its precise model is usually hard to be constructed using classical analytical approaches [12–18]. Thus, data-driven approaches are utilized for process modeling and fault isolation. However, when applied to different power plants, the HFPS yields very different input/output functions due to the change of its loads and working environments. This means that there is no general process model for HFPS in all conditions.

For each HFPS that has been installed, a particular process model should be constructed. Therefore, for conventional SVR-based fault isolation, experienced operators are required to select the modeling parameters at site.

In this section, F-SVR was applied in order to isolate the fault of the high frequency power supply. Three operating cases were investigated: normal status, overcurrent fault, and learning excitation fault. Data for the 3 operating cases of HFPS was prepared in Figure 10 and an overcurrent fault sample was used as the input data for testing in Figure 11.

The basic scheme for HFPS fault isolation is designed in Figure 12. Firstly, process data for each operating case is acquired; then, F-SVR is implemented to approximate the unknown function between input (time) and output (the first primary current) for each operating case. As the models for all operating cases have been established, the residual vectors can be generated and then by finding the model with the smallest residual the fault can be isolated.

In this experiment, F-SVR is implemented to approximate the functions between the input and the output, namely, $G_1(x)$, $G_2(x)$, and $G_3(x)$. As shown in Figure 13, the functions we obtained using F-SVR accurately describe the relationship between the input and the output of the unknown functions for the different operating cases without setting parameters.

Since $G_1(x)$, $G_2(x)$, and $G_3(x)$ were already obtained by the F-SVR method, the residuals could be generated
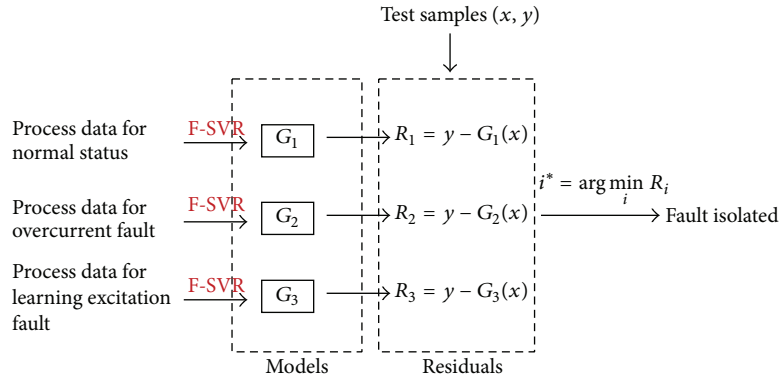
FIGURE 12: Fault isolation scheme for HFPS.

(a) Function approximation for normal status: $G_1(x)$

(b) Function approximation for overcurrent fault: $G_2(x)$

(c) Function approximation for leaning excitation fault: $G_3(x)$
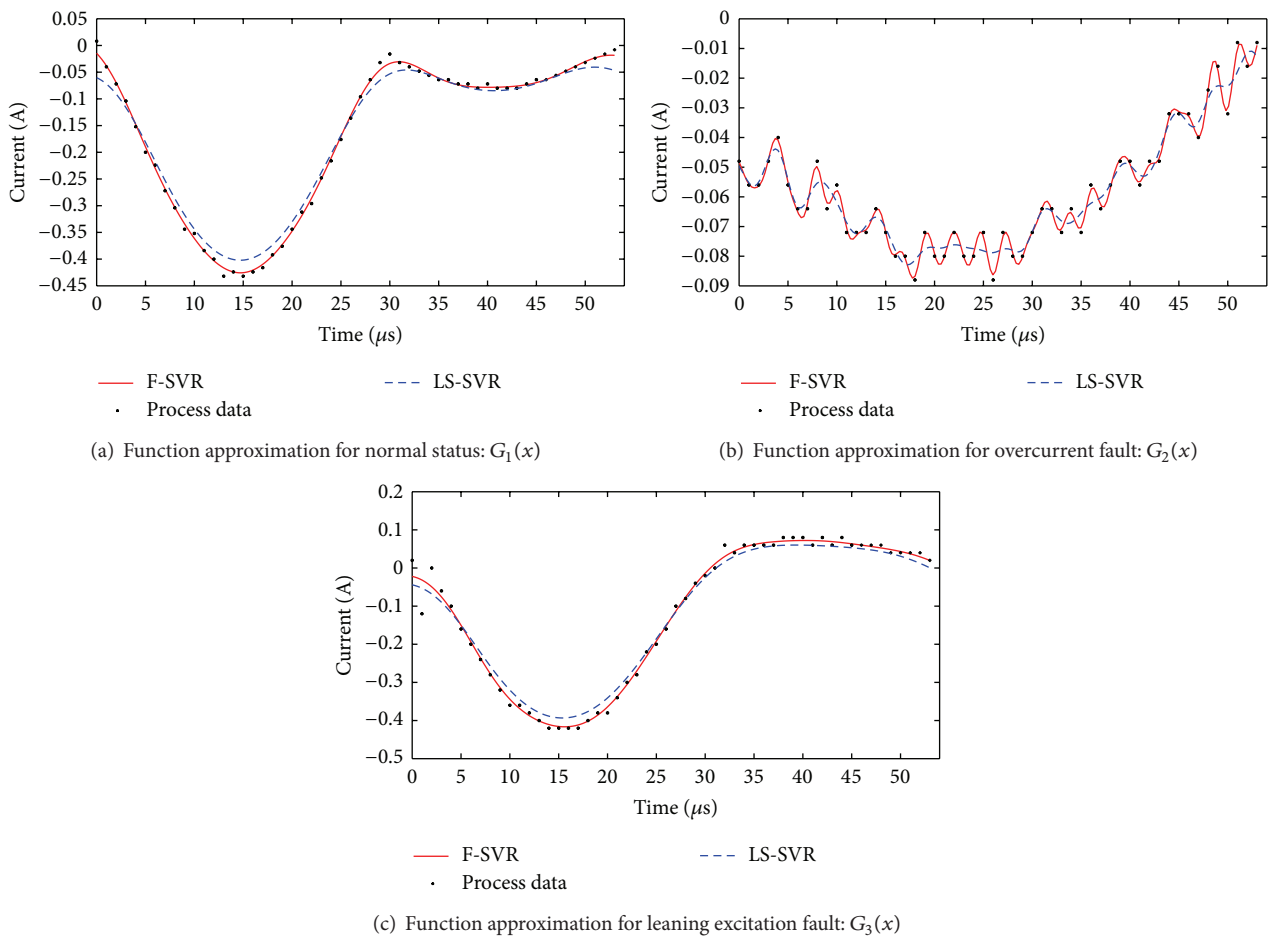
FIGURE 13: Function approximation using F-SVR and LS-SVR.

using (12). The corresponding residuals are recorded in Table 1. According to Table 1, the testing sample belongs to the overcurrent fault ($G_2(x)$ yields the smallest residual). Based on prior knowledge of the testing sample, the diagnostic result is consistent and shows the feasibility of our method.

The LS-SVR method [11, 19] is also implemented to give a comparison. As is shown in Table 1, the LS-SVR method

also makes a correct diagnosis and it has a better ability to generalize than our method (the number of SVs is smaller). However, our method yields a smaller value of residuals. This implies that our method produces a better modeling accuracy. Most importantly, all parameters are required to be selected manually in LS-SVR but are selected automatically in F-SVR.

TABLE 1: Diagnostic performance of F-SVR and LS-SVR.

Operating status	Residuals (R_i)		Number of SVs (d)		Parameters setting	
	F-SVR	LS-SVR	F-SVR	LS-SVR	F-SVR	LS-SVR
Normal (G_1)	15.912	16.231	26	23	Auto	Manual
Overcurrent (G_2)	**11.296**	**14.131**	38	28	Auto	Manual
Learning excitation (G_3)	17.104	17.932	23	20	Auto	Manual

5. Conclusions

SVR is one of the most efficient tools for fault diagnosis because it is able to accurately model a function between the input and the output using process data. However, even though the SVR approach has been utilized for over a decade, there is still no consensus within the community on how to adequately select regression parameters. Given that the F-SVR method offers an automatic selection for regression parameters, we chose it to implement the fault isolation for nonlinear systems. We demonstrated on both a numerical experiment and the fault isolation for HFPS that F-SVR is especially suited for cases that yield complicated sample distributions. This is because this method generates reasonable parameters for each region by dividing the training samples into different regions according to the sample distribution complexity. Based on this work, we hope that F-SVR will become more widely recognized as a preferred fault isolation for nonlinear systems.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] S. X. Ding, *Model-Based Fault Diagnosis Techniques: Design Schemes, Algorithms, and Tools*, Springer, Berlin, Germany, 2008.

[2] Z. N. Sadough Vanini, K. Khorasani, and N. Meskin, "Fault detection and isolation of a dual spool gas turbine engine using dynamic neural networks and multiple model approach," *Information Sciences*, vol. 259, pp. 234–251, 2014.

[3] F. Filippetti, G. Franceschini, C. Tassoni, and P. Vas, "Recent developments of induction motor drives fault diagnosis using AI techniques," *IEEE Transactions on Industrial Electronics*, vol. 47, no. 5, pp. 994–1004, 2000.

[4] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 1995.

[5] B. Ravikumar, D. Thukaram, and H. P. Khincha, "Application of support vector machines for fault diagnosis in power transmission system," *IET Generation, Transmission and Distribution*, vol. 2, no. 1, pp. 119–130, 2008.

[6] A. A. Yusuff, C. Fei, A. A. Jimoh, and J. L. Munda, "Fault location in a series compensated transmission line based on wavelet packet decomposition and support vector regression," *Electric Power Systems Research*, vol. 81, no. 7, pp. 1258–1265, 2011.

[7] B. Samanta, K. R. Al-Balushi, and S. A. Al-Araimi, "Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection," *Engineering Applications of Artificial Intelligence*, vol. 16, no. 7-8, pp. 657–665, 2003.

[8] H. Yi, X. F. Song, B. Jiang, Y. F. Liu, and Z. H. Zhou, "Flexible support vector regression and its application to fault detection," *Acta Automatica*, vol. 39, no. 3, pp. 272–284, 2013.

[9] J. R. Bunch and L. C. Kaufman, *A Computational Method of Indefinite Quadratic Programming Problem*, Linear Algebra and Its Applications, Elsevier-North Holland, 1980.

[10] V. Cherkassky and Y. Ma, "Practical selection of SVM parameters and noise estimation for SVM regression," *Neural Networks*, vol. 17, no. 1, pp. 113–126, 2004.

[11] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999.

[12] X.-Z. Gao, H. Xu, X. Wang, and K. Zenger, "A study of negative selection algorithm-based motor fault detection and diagnosis," *International Journal of Innovative Computing, Information and Control*, vol. 9, no. 2, pp. 875–901, 2013.

[13] R. Qi, L. Zhu, and B. Jiang, "Fault-tolerant reconfigurable control for MIMO systems using online fuzzy identification," *International Journal of Innovative Computing, Information and Control*, vol. 9, no. 10, pp. 3915–3928, 2013.

[14] X. Su, P. Shi, L. Wu, and S. Nguang, "Induced L2 filtering of fuzzy stochastic systems with time-varying delays," *IEEE Transactions on Cybernetics*, vol. 43, no. 4, pp. 1251–1264, 2013.

[15] D. Xu, B. Jiang, H. Liu, and P. Shi, "Decentralized asymptotic fault tolerant control of near space vehicle with high order actuator dynamics," *Journal of the Franklin Institute: Engineering and Applied Mathematics*, vol. 350, no. 9, pp. 2519–2534, 2013.

[16] X. Su, L. Wu, and P. Shi, "Senor networks with random link failures: distributed filtering for T-S fuzzy systems," *IEEE Trans on Industrial Informatics*, vol. 9, no. 3, pp. 1739–1750, 2013.

[17] O. Inan, M. Serter Uzer, and N. Yilmaz, "A new hybrid feature selection method based on association rules and PCA for detection of breast cancer," *International Journal of Innovative Computing, Information and Control*, vol. 9, no. 2, pp. 727–729, 2013.

[18] B. Jiang, M. Staroswiecki, and V. Cocquempot, "Fault accom-
     modation for nonlinear dynamic systems," *Institute of Electrical
     and Electronics Engineers. Transactions on Automatic Control*,
     vol. 51, no. 9, pp. 1578–1583, 2006.

[19] Y. Guo, C. Ran, X. Li, and J. Ms, "Adaptive online prediction
     method based on LS-SVR and its application in an electronic
     system," *Journal of Zhejiang University-SCIENCE C, (Computers
     & Electronics)*, vol. 13, no. 12, pp. 881–890, 2012.