A New Time Series Prediction Method Based on LS-SVR and AGO

Guo Yangming, Zhang Lu, Li Xiaolei, Ran Congbao, Ma Jiezhong

School of Computer Science and Technology, Northwestern Polytechnical University, Xi’an, China, yangming_g@nwpu.edu.cn

Abstract

Fault or health condition prediction of complex systems has attracted more and more attention in recent years. Because it is difficult to establish precise physical model, the time series of complex systems are often used to implement the prediction in practice, and LS-SVR is widely used in time series prediction. In order to improve the prediction accuracy, based on grey system theory, accumulated generating operation (AGO) with raw time series is made to improve the data quality and regularity, and then inverse accumulated generating operation (IAGO) is performed to get the prediction results. In addition, because the appropriate kernel function has great impact with the prediction accuracy with LS-SVR, we propose a modified Gaussian RBF that satisfies the two requirements of distance functions-based kernel functions, which are fast damping at the place adjacent to the test point and moderate damping at infinity. The results indicate preliminarily that the new method has better performance and it is an effective prediction method for its good prediction precision.

Keywords: Least Squares Support Vector Regression (LS-SVR), Accumulated Generating Operation (AGO), Gaussian RBF, Time Series Prediction

1. Introduction

Fault or health trend prediction is essential to ensure the operational safety of complex systems. However, the complex systems are usually non-linear systems, and it is difficult to establish their accurate mathematical models. So in practice, time series-based prediction models have attracted increased attention, such as Artificial Neural Networks (ANN), Support Vector Regression (SVR), etc. [1-6]

Comparing with SVR, ANN has a number of shortcomings, such as it always has long training periods, easily falls into local minimum traps, hard determines the hidden layer size and learning rate and has poor generalization capacity [7,8]. Based on statistical theory and structural risk minimization principle [9], SVR can overcome the problems in ANN. SVR has a global optimum and exhibits better accuracy in nonlinear and non-stationary time series data prediction via kernel function. However, aiming at the large sample data, SVR solving quadratic programming (QP) problem is more complex. So, LS-SVR was proposed by Suykens et al.[10] In LS-SVR, the inequality constrains were replaced by equality constrains which reduces the calculation cost effectively. LS-SVR has more attention in time series forecasting. [11-15]

However in actual application, the raw time series always show the incompleteness, randomness and inconsistency, which will reduce the prediction accuracy. Therefore, some researchers proposed to combine several methods to improve prediction effect[16,17], which can fusion their advantages and avoid their drawbacks. In this paper, based on the grey system theory[18], we propose a new combination time series prediction method to obtain better accuracy. Here, we make accumulated generating operation (AGO) with raw time series to reduce the randomness and to improve the data regularity and quality, then we do the prediction using LS-SVR, and finally we can obtain the results by inverse accumulated generating operation (IAGO).

Moreover, there is another key issue for LS-SVR that is to select appropriate kernel function and its parameters. It has a great impact on the model accuracy and generalization capabilities. Up to now, there are some most widely used kernel functions, which include the radial basis function (RBF) kernel, polynomial kernel, sigmoid kernel and linear kernel etc. The polynomial kernel and RBF kernel always satisfies Mercer’s theorem, whereas other kernels satisfy this theorem only for certain conditions[19].
In comparison with above feasible kernel functions, the most commonly adopted one is the RBF kernel, named Gaussian RBF. As a distance function-based kernel function, Gaussian RBF is a more compactly supported kernel function, which can reduce the computational complexity of training process, improve generalization capability and accuracy of LS-SVR. So Gaussian RBF is always selected as kernel function in time series prediction.

However, according to the characters of distance function-based kernel function, closely test points in original space will become extremely sparseness after mapping to the high dimensional space by Gaussian RBF[20]. In order to overcome this drawback, the Gaussian RBF must be modified to satisfy two requirements: one is that the function has fast damping at the place adjacent to the test point, the other one is that the function can keep a moderate damping at infinity[20]. The Gaussian RBF only satisfies the fist condition. In this paper, we present a new modified Gaussian RBF that will both satisfy the two conditions. And then we use LS-SVR with this modified Gaussian RBF to perform time series prediction.

The remainder of the paper will be structured as follows: section 2 gives a brief introduction to LS-SVR; Section 3 proposes the prediction method with time series based on AGO, LS-SVR and IAGO; Section 4 presents the form of the new modified Gaussian RBF; Section 5 shows simulation experiments; and Section 6 gives the conclusions.

2. Least squares support vector regression

Consider a training sample data set \( \{ (x_k, y_k) \} \) with input data \( x_k \in \mathbb{R}^n \) and output \( y_k \in \mathbb{R} \), where \( n \) denotes the number of training samples. The goal of LS-SVR is to obtain a function as follows

\[
y(x) = w^T \phi(x) + b
\]

where the nonlinear mapping function \( \phi(\cdot) \) maps the input data into a higher dimensional feature space. That means the method makes the nonlinear fitting problem in input feature space to be replaced by a high-dimensional feature space linear fitting problem. \( w \) is the weight vector and \( b \) is bias constant.

According to the structure risk minimization principle, using the training sample data, \( w \) and \( b \) can be found through a constrained convex optimization problem as follows

\[
\min J(w, e) = \frac{1}{2} w^T w + \frac{1}{2} c \sum_{i=1}^{n} e_i^2
\]

s.t. \( y_k = w^T \phi(x_k) + b + e_k, \quad k = 1,2,\ldots,N \)

where \( J \) is loss function, \( e_k \in \mathbb{R} \) is the slack variables, \( c \) is a regularization parameter. By transforming this optimization problem into dual form with Lagrange function as follows

\[
L(w, b, e, \alpha) = J(w, e) - \sum_{i=1}^{n} \alpha_i \left( w^T \phi(x_i) + b + e_i - y_i \right)
\]

where \( \alpha_i \) are the Lagrange multipliers.

It is obvious that the optimal solution of Eq.(2) satisfies Karush-Kuhn-Tucker (KKT), then the optimal conditions are shown as follows

\[
\begin{align*}
\frac{\partial L}{\partial w} &= w - \sum_{i=1}^{n} \alpha_i \phi(x_i) = 0 \Rightarrow w = \sum_{i=1}^{n} \alpha_i \phi(x_i) \\
\frac{\partial L}{\partial b} &= -\sum_{i=1}^{n} \alpha_i = 0 \Rightarrow \sum_{i=1}^{n} \alpha_i = 0 \\
\frac{\partial L}{\partial \alpha_i} &= w^T \phi(x_i) + b + e_i - y_i = 0 \Rightarrow y_i = w^T \phi(x_i) + b + e_i \\
\frac{\partial L}{\partial e_i} &= c e_i - \alpha_i = 0 \Rightarrow e_i = \frac{1}{c} \alpha_i
\end{align*}
\]

\[ \Box \]
After eliminating $w$ and $e_i$, we could have the solution by following linear equations

$$
\begin{bmatrix}
0 & 1^n
\end{bmatrix}
\begin{bmatrix}
\mathbf{a} \\
\mathbf{b}
\end{bmatrix} =
\begin{bmatrix}
0
\end{bmatrix}
$$

(5)

where $\mathbf{K}(i,j) = k(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$, $\mathbf{a} = [\alpha_1, \alpha_2, \ldots, \alpha_n]^T$, $1_n$ is $n$-dimensional vector of all ones, $\mathbf{I}$ is a unite matrix and $\mathbf{y} = [y_1, y_2, \ldots, y_n]^T$. Saddle point system Eq.(5) can be factorized into a positive definite system[21].

In LS-SVR, the optimization problem is simplified to solve linear equations instead of more complex quadratic programming problem in SVR. Therefore, the computational complexity is decreased significantly. The system of linear Eq.(5) is nonsingular, from which the solutions of $\alpha_i$ and $b$ could be obtained. Hence, the LS-SVR model for function estimator can be expressed as follows

$$y(x) = \sum_{i=1}^{n} \alpha_i \varphi(x_i, x) + b$$

(6)

### 3. AGO-based prediction method for time series

The good precision of LS-SVR always depends on the high quality of training sample data. As the core of grey prediction theory, accumulated generating operation (AGO) has a main advantage that it can reduce the disturbance with the stochastic factors. AGO makes the rules hidden the raw time series to be presented fully and enhance the law of the data.

Consider the raw time series $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n))$, and take the AGO with $x^{(0)}$ as follows

$$x^{(1)}(i) = \sum_{j=1}^{i} x^{(0)}(j), \quad i = 1, 2, \ldots, n$$

(7)

The $x^{(1)}(1), x^{(1)}(2), \ldots, x^{(1)}(n)$ will be new time series.

After new time series $x^{(1)}$ is formed, it can be used to predict the future values based on the previous and current values. The previous and the current values of the time series are used as input for the prediction model

$$\{x^{(1)}(t+1), x^{(1)}(t+2), \ldots, x^{(1)}(t+h)\} = F(x^{(1)}(t), x^{(1)}(t-1), \ldots, x^{(1)}(t-m+1))$$

(8)

where $h$ represents the number of ahead predictions, $F$ is prediction model and $m$ is size of the repressor. According to Eq.(8), we can obtain the training sample.

In this paper, we use a prediction strategy that the predicted values as known data to predict the next one. The model can be constructed by making one-step prediction

$$x^{(1)}(t+1) = F(x^{(1)}(t), x^{(1)}(t-1), \ldots, x^{(1)}(t-m+1))$$

(9)

The regressor of the model is defined as the vector of inputs $(x(t), x(t-1), \ldots, x(t-m+1))$.

After get the prediction sequence, we can compute the prediction results by inverse accumulated generating operation (IAGO)

$$\hat{x}^{(0)}(t+1) = \hat{x}^{(1)}(t+1) - x^{(1)}(t)$$

(10)

where $\hat{x}^{(0)}(t+1)$ and $x^{(1)}(t+1)$ are respectively the one-step prediction values of the raw time series and the time series via AGO.

### 4. Modified Gaussian RBF

After establishing the LS-SVR model, the key issue is how to select an appropriate kernel function which greatly influences the performance of LS-SVR. In practice, RBF is always selected as kernel function with a strong learning ability. It can realize nonlinear mapping, reduce
computational complexity and improve the generalization performance. As a result, the stander Gaussian RBF that we employ as kernel function in this paper is presented as following

$$k(x, x_i) = \exp\left(\frac{-\|x - x_i\|^2}{2\sigma^2}\right)$$  \hspace{1cm} (11)

where \(\sigma\) is the width of kernel and it is a positive real constant.

Based on the discussion of RBF on section 1, although the stander Gaussian RBF has some shortages, it is still a good kernel function. So some researchers proposed some scheme to improve the performance of the stander Gaussian RBF[22,23].

Although it is obviously from Eq.(11) that \(\sigma\) is the only one parameter. The research results of [22] indicate that the adjustment of \(\sigma\) can’t achieve good effect, i.e., it can’t improve the inherent in the stander Gaussian kernel too much. In [23], Huang X. proposed a new kernel function (shown below) that distance equation serves as denominator of exponential functions

$$k(x, x_i) = \frac{\sigma^2}{\left(1 - \frac{\|x - x_i\|^2}{\sigma^2}\right)^{\chi + \kappa}}$$  \hspace{1cm} (12)

It shows from the characteristic curve of Eq.(12) that Eq.(12) satisfies the two conditions better than the stander Gaussian RBF.

According to the ideal of Eq.(12), we propose an improved kernel function to achieve better performance. Its form is shown as follows

$$k(x, x_i) = \sigma^2\left(1 - \frac{\|x - x_i\|^2}{\sigma^2}\right)^{\chi + \kappa}$$  \hspace{1cm} (13)

where \(\chi\) is displacement parameter and \(\kappa\) is amplitude parameter. They are used to adjust the displacement and amplitude of kernel function. The Figure 1 shows the comparison of the stander Gaussian RBF and improved Gaussian RBF on damping with different distance from the test point.

![Damping Comparison of Gaussian RBF and Modified RBF](image)

**Figure 1.** Damping comparison of Gaussian RBF and modified RBF

From Fig.1, it is obvious that the modified Gaussian RBF can satisfy the two conditions above. Especially, the modified kernel function far away from test point can keep some attenuation and is not almost zero. And for this modified Gaussian RBF, we can perform joint optimization with all parameters to get best kernel function. Hence, the modified Gaussian RBF is an advisable kernel function with improving the performance of kernel function.

### 5. Simulation experiments and results

In this section, we conduct three experiments to demonstrate the performance of the proposed method on prediction accuracy. All the experiments adopt MatlabR2011b with LS-SVMLab1.8 Toolbox.
5.1. Estimation of the prediction method

If the prediction value has less relative error, the LS-SVR model is a better model. Root Mean Squared Error (RMSE) [24] of prediction is usually as the evaluation criteria. It is defined as follows

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x(k) - \hat{x}(k))^2}
\]  

(14)

where \( n \) is the number of training sample data, \( \hat{x}(k) \) and \( x(k) \) are the prediction value and actual value respectively.

5.2. Experiment I and analysis

In order to validate the performance of the proposed AGO-based method, we performed Experiment I with 75 sample time series data of variable \( x \), which come from one complex avionics equipment (shown in Figure. 2).

![Figure 2. Raw time series of avionics equipment](image)

We set the front 50 time series data as training samples, and any continuous 6 are taken as a sample, where the data of the first 5 data compose an input sample vector and last one is as the output vector, i.e., in the example, we have 45 training data. And then we predict the No.51 to No.75 time series data using the trained model. To avoid over fitting, the experiment is repeated 100 times and the average results are obtained.

In the Experiment I, we compare the proposed method (called ALS-SVR) with traditional LS-SVR presented in [4] on prediction accuracy. Here, Gaussian RBF kernel \( k(x, y) = \exp\left(-\frac{||x-y||^2}{2\sigma^2}\right) \) is adopted as kernel function, and the parameters are jointly optimized with traditional gridding search method, where the search rang for \( c \) and \( \sigma^2 \) is \([10, 3000]\). The prediction results are shown in Figure 3 and the prediction RMSE are reported in Table 1.
Table 1. Prediction RMSE

<table>
<thead>
<tr>
<th>Method</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional LS-SVR</td>
<td>2.4411</td>
</tr>
<tr>
<td>ALS-SVR</td>
<td>2.1569</td>
</tr>
</tbody>
</table>

From Table 1 and Figure 3, we can see that the proposed method (ALS-SVR) has better prediction accuracy since ALS-SVR effectively avoid the random disturbances existing in the raw time series and improve the regularity of the time series.

5.3. Experiment II and analysis

In order to evaluate the proposed modified Gaussian RBF, we perform simulation Experiment II with \( \sin x / x \) defined on the interval \([0.1, 40]\) with step 0.1. In this experiment, we compared the prediction accuracy via LS-SVR with the stander Gaussian RBF and the modified Gaussian RBF respectively. Moreover, we added Gaussian noise with noise std. 0.05 into the simulation time series, i.e., the experiment is implemented on a random set of samples corrupted by additive zero-mean Gaussian noise (shown in Figure 4).
We select 400 time series data as samples and set the front 200 time series data as training samples. We take the data of any continuous 11 as a sample, where the data of the first 10 data compose an input sample vector and last one as the output vector, i.e., in the simulation example, we have 190 training data. And then we predict the No. 201 to No. 200 time series data using the trained model. We run 100 times of prediction and the average of them will be taken.

In Experiment II, the parameters are set with $\sigma = 0.3$, $\kappa = 3$, and $\chi = 0.1$. The prediction results are shown in Figure 5 and Table 2.

![Prediction Results graph]

**Figure 5.** Prediction results

<table>
<thead>
<tr>
<th>Method</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>LS-SVR with RBF</td>
<td>0.5342</td>
</tr>
<tr>
<td>LS-SVR with modified</td>
<td>0.3505</td>
</tr>
<tr>
<td>Gaussian RBF</td>
<td></td>
</tr>
</tbody>
</table>

As shown in Figure 5 and Table 2, the LS-SVR with modified Gaussian RBF has better prediction results. That means the information is better presented with the modified Gaussian RBF in mapping feature space.

**5.4. Experiment III and analysis**

In Experiment III, we adopt an application of a certain complex system to demonstrate the accuracy of the proposed prediction method using modified RBF and AGO.

We collect 70 data from the complex system (shown in Figure 6, omitted dimension), and use the front 50 group as training samples, and then we predict the last 20 data. Same as Experiment I and II, we also run 100 times of prediction and take the average of them. In addition, all the parameters of modified RBF will be jointly optimized with traditional gridding search method, where the search range is $[0.01, 500]$.

The prediction results (shown in Figure 7 and Figure 8) indicate the proposed prediction method is a good approach in practice for its higher accuracy. In fact, the mean prediction RMSE is 0.0879.
A New Time Series Prediction Method Based on LS-SVR and AGO
Guo Yangming, Zhang Lu, Li Xiaolei, Ran Congbao, Ma Jiezhong

Figure 6. Raw data

Figure 7. Prediction results

Figure 8. Prediction error
6. Conclusions

In this work, we address a new prediction method with LS-SVR to improve time series prediction accuracy. The main contributions of the method are that we make the accumulated generating operation (AGO) with raw time series to reform the quality of raw time series firstly; and then we aim at the shortage of the stander Gaussian RBF and modify it. The new modified Gaussian RBF can satisfy the two requirements of distance function-based kernel function, which are fast damping at the place adjacent to the test point and moderate damping at infinity. We conduct assessment experiments that the results demonstrate the proposed prediction method is a better approach and can improve prediction accuracy effectively in practice.

7. Acknowledgements

This work is supported by Major Project of Chinese National Programs for Fundamental Research and Development (973 Program), National Natural Science Foundation (NNSF) of China under Grant No. 61001023 and No. 61101004, Basic Research Program of Shaanxi Province under Grant No. 2010JQ8005 and Aviation Science Fund of China under Grant No. 2010ZDS039. And the authors would like to thank Reliability Research Laboratory of University of Alberta for the experiment support.

8. References


