Time Series Prediction Method Based on LS-SVR with Modified Gaussian RBF

Yangming Guo[1,*], Xiaolei Li[1], Guanghan Bai[2], and Jiezhong Ma[1]

[1] School of Computer Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China
[2] Department of Mechanical Engineering, University of Alberta, Edmonton T6G2G8, Canada
{yangming_g,majz}@nwpu.edu.cn, 465628547@qq.com, guanghan@ualberta.ca

Abstract. LS-SVR is widely used in time series prediction. For LS-SVR, the selection of appropriate kernel function is a key issue, which has a great impact with the prediction accuracy. Compared with some other feasible kernel functions, Gaussian RBF is always selected as kernel function due to its good features. As a distance functions-based kernel function, Gaussian RBF also has some drawbacks. In this paper, we modified the standard Gaussian RBF to satisfy the two requirements of distance functions-based kernel functions which are fast damping at the place adjacent to the test point and keeping a moderate damping at infinity. The simulation results indicate preliminarily that the modified Gaussian RBF has better performance and can improve the prediction accuracy with LS-SVR.

Keywords: Least squares support vector regression (LS-SVR), Gaussian RBF, Time series prediction.

1 Introduction

Fault or health trend prediction is essential to ensure the safe operation of complex equipments. However, the complex equipments are usually nonlinear systems. It is difficult to establish their accurate mathematical models. So in practice, the observation time series are often used to implement the prediction of complex equipments. Many nonlinear prediction methods are presented for time series prediction.[1-7] Support Vector Machine (SVM) based methods are one of the widely applied methods.

SVM was firstly proposed by Vapnik et al.[8] in mid-nineties. It is a machine learning method based on the statistical learning theory. SVM has strict mathematical basis, which improves the generalization ability by employing structural risk minimization principle and provides better solutions to many practical problems, such as the problem with small sample high dimensional data, nonlinear problem and local minima problem, etc.

The standard SVM solves a convex quadratic optimization problem where solution is unique; however the solving process is complicated. Suykens et al.[9] presented

T. Huang et al. (Eds.): ICONIP 2012, Part II, LNCS 7664, pp. 9–17, 2012.

least squares support vector machine (LS-SVM) via transferring the inequality constraints into equality constraints and replacing the empirical risk deviation by quadratic deviation. LS-SVM retains the advantages of SVM in solving the small sample size data and nonlinear problems. LS-SVM also has faster solving speed and requires less computing resource. Later, Vapnik et al.[8] extended LS-SVM to the regression estimation of nonlinear systems, called LS-SVR[4]. However, for LS-SVR and the other SVM-based methods, it is very important to select an appropriate kernel function and its parameters. This key issue has great impact on the model accuracy and generalization capability.

There are some most widely used kernel functions which include radial basis function (RBF) kernel, polynomial kernel, sigmoid kernel and linear kernel etc. Polynomial kernel and RBF kernel always satisfies Mercer's theorem, the other kernels satisfy Mercer's theorem only under some certain conditions[10].Moreover, polynomial kernel has more parameters than RBF kernel. Thus, in comparison with above feasible kernel functions, the most commonly adopted RBF kernel is 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 and improve the generalization performance 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 functions, the closeness test points in original space will become extremely sparseness after being mapped to the high dimensional space by Gaussian RBF, i.e., Gaussian RBF kernel is a local kernel with a stronger learning ability but weaker dissemination ability.[11,12] In order to overcome this drawback, 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 is that the function can keep a moderate damping at infinity[13]. 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 apply LS-SVR with this modified Gaussian RBF to perform time series prediction.

The remainder of the paper will be structured as follows: Section 2 proposes the form of the new modified Gaussian RBF; Section 3 shows two experiments and analysis; and the conclusions are drown in Section 4.

2 Modified Gaussian RBF

Consider a training sample data set $\{(x_k, y_k)\}_{k=1}^{N}$ with input data $x_k \in R^n$ and output $y_k \in R$, where N denotes the number of training samples. The LS-SVR model for function estimator can be expressed as follows

$$y(x) = \sum_{k=1}^{N} \alpha_k \varphi(x)^T \varphi(x_k) + b = \sum_{k=1}^{N} \alpha_k K(x, x_k) + b \tag{1}$$

where b is the bias constant, α_k are the Lagrange multipliers, 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. $K(x, x_k) = \varphi(x)^T \varphi(x_k)$ is the kernel function. The kernel function can perform a nonlinear mapping to a high-dimensional feature space by replacing the inner product for nonlinear pattern problem.

After establishing the LS-SVR model, the key issue is how to select an appropriate kernel function which greatly influences the performance of the LS-SVR model. In practice, RBF is always selected as kernel function for its 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 follows

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

where σ is the width of kernel and it is a positive real constant.

Based on the discussion of RBF in section 1, although the standard Gaussian RBF has some shortages, it is still a good kernel function. So some researchers proposed some schemes to improve the performance of the standard Gaussian RBF[14,15].

Although it is obviously from Eq.(2) that σ is the only one parameter, the research results of Ref.[14] indicated that the adjustment of σ can't achieve more better effect. Ref.[15] proposed a new kernel function, in which the distance equation served as the denominator of the exponential function, shown as follows

$$k(x, x_i) = 2^{\left(\frac{\sigma^2}{\|x - x_i\|^2}\right)} \tag{3}$$

From the characteristic curve of Eq.(3), we can see that Eq.(3) is better able to satisfy the two conditions (expressed in section 1) than the standard Gaussian RBF.

According to the idea of Eq.(3), we modify the standard Gaussian RBF to achieve better performance. Its form is shown as follows

$$k(x, x_i) = \kappa^{\left(\frac{\sigma^2}{\|x - x_i\|^2 + \chi}\right)} \tag{4}$$

where, we call χ as displacement parameter and κ as amplitude parameter. They are used to adjust the displacement and amplitude of kernel function. In order to simple the selection complexity, Eq.(4) can be transferred with the following form

$$k(x, x_i) = e^{(\frac{\sigma^2}{\|x - x_i\|^2 + \chi}) \ln \kappa} \tag{5}$$

For the description consistency with above equations, rewrite Eq.(5) as follows

$$k(x, x_i) = e^{(\frac{\sigma^2}{\|x - x_i\|^2 + \chi})} \tag{6}$$

where $\sigma^2 \ln \kappa \rightarrow \sigma^2$.

Fig.1 shows the comparison of the standard Gaussian RBF and the modified Gaussian RBF (see Eq.(6)) on damping with different distance from the test point.

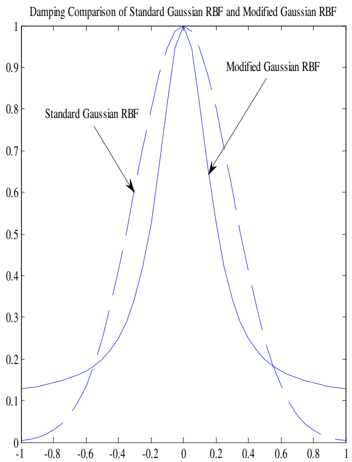

Fig. 1. Damping comparison of standard and modified Gaussian RBF

From Fig.1, it is obvious that the modified Gaussian RBF can satisfy the two conditions described in section 1. Especially, the modified Gaussian RBF can keep some attenuation and does not almost equal to zero at points far away from the test point. For this modified Gaussian RBF, we can perform joint optimization with all parameters to get the best kernel function. Hence, the modified Gaussian RBF is an advisable kernel function due to the improved performance of kernel function.

3 Time Series Prediction Experiments and Analysis

In this section, we perform time series prediction simulation and application experiments to evaluate the modified Gaussian RBF. All the experiments adopt MatlabR2011b and LS-SVMlab1.8 Toolbox (The software can be downloaded from http://www.esat.kuleuven.be/sista/lssvmlab) under Windows XP operating system to get the experimental results. To avoid over fitting, all the experiments are run 100 times and the average values are taken.

3.1 Prediction Estimation Criterion

If the prediction value has low relative error, the LS-SVR model is a better model. The root mean squared error (RMSE) [16] of prediction is usually as evaluation criterion. The prediction RMSE is defined as follows

$$RMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(x(k) - x'(k))^2} \tag{7}$$

where n is the number of training sample data, $x'(k)$ and $x(k)$ are the prediction and actual values respectively.

3.2 Simulation Experiment and Analysis

In the simulation experiment, we apply $\sin x / x$ defined on the interval [0.1, 40] with step 0.1 to generate sample time series. In this simulation experiment, we compare the prediction accuracy and time cost via LS-SVR with the standard Gaussian RBF and the modified Gaussian RBF respectively. Moreover, we add 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 Fig.2).

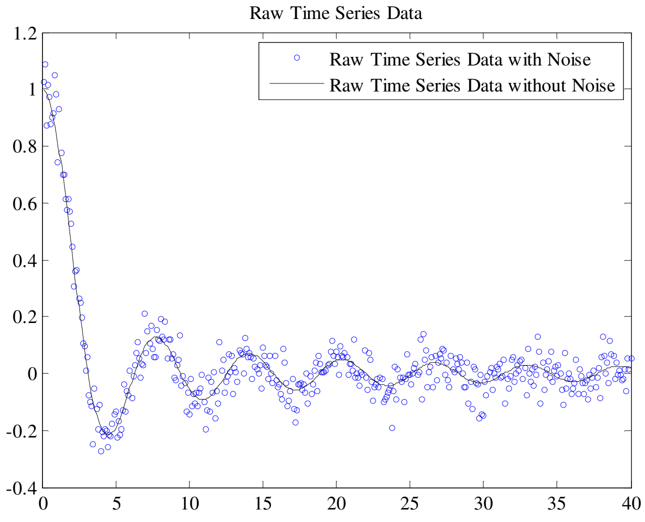

Fig. 2. Raw time series data of simulation experiment

In practice, researchers often use part of the sample data to train the LS-SVR model and use another part of the sample data to validate. So we select 400 time series data as samples and set the first 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.

In the simulation experiment, all the parameters are selected by traditional grid search method, where the search rang is [0.01, 200]. The prediction results, which are training time(TrTime) , prediction time (PrTime) and RMSE, are shown in Fig.3 and Table 1.

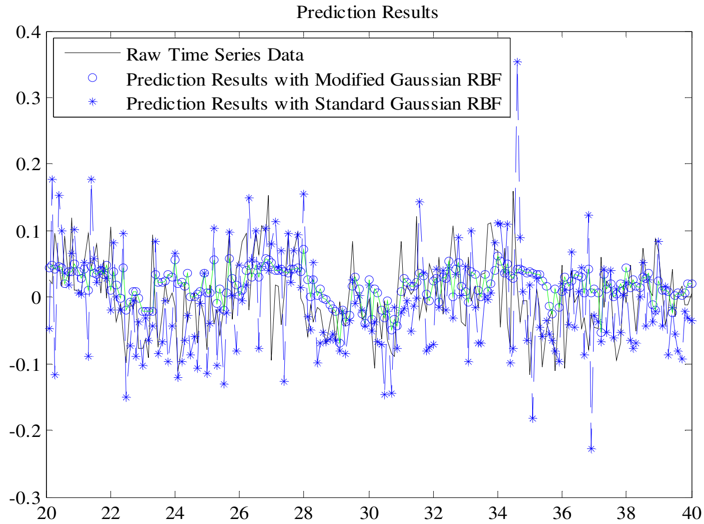

Fig. 3. Prediction results of simulation experiment

Table 1. Prediction results of simulation experiment

Method	TrTime/s	PrTime/s	RMSE
Standard Gaussian RBF	0.012309	0.014164	0.5342
Modified Gaussian RBF	0.011180	0.004972	0.3505

From the Fig.3 and Table 1, we can see that the LS-SVR with modified Gaussian RBF has better prediction results on computing time and RMSE. This may be due to that the information in sample data is better presented by the modified Gaussian RBF in the high-dimensional mapping feature space.

3.3 Application Experiment and Analysis

In application experiment, we collect 60 data from a complex system (shown in Fig.4). we use the first 50 group as training samples, and then predict the last 10 data. All the parameters of modified RBF will be jointly optimized via the traditional grid search method with the search rang [0.01, 200].

The prediction results (shown in Fig.5 and Fig.6) indicate that the modified Gaussian RBF can improve the performance of LS-SVR, and the modified Gaussian RBF is a good and effective kernel function in practice.

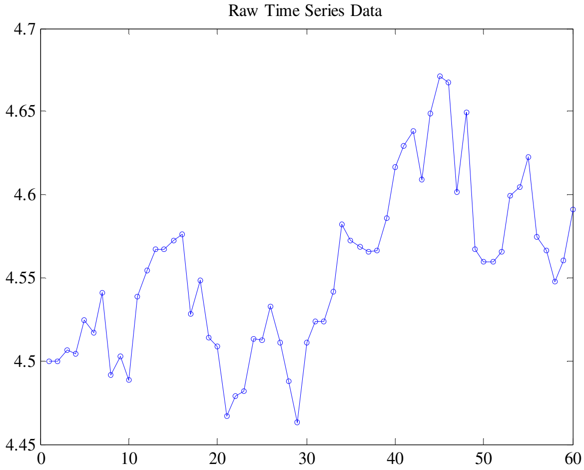

Fig. 4. Raw time series data of application experiment

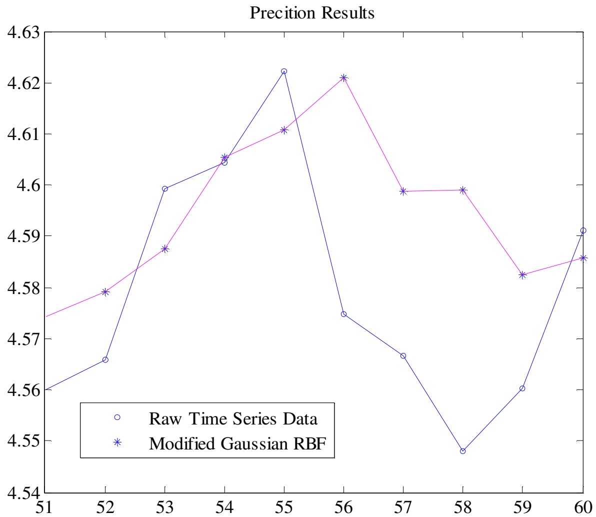

Fig. 5. Prediction results of application experiment

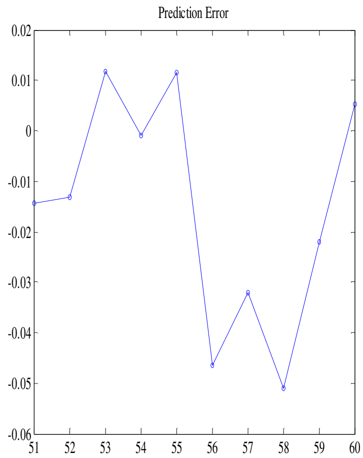

Fig. 6. Prediction error of application experiment

4 Conclusions

In this paper, we aim at the shortages of standard Gaussian RBF and modify it. The modified Gaussian RBF can satisfy the two requirements of distance function-based kernel functions, which can fast damping at the place adjacent to the test point and keep a moderate damping at infinity. We apply LS-SVR with the modified Gaussian RBF to perform time series prediction experiments. The results indicate that the modified Gaussian RBF has better performance and can improve prediction accuracy effectively.

Acknowledgments. The 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. 2010ZD53039. The authors also gratefully acknowledge the support from the Reliability Research Laboratory of University of Alberta.

References

1. Müller, K.-R., Smola, A.J., Rätsch, G., Schölkopf, B., Kohlmorgen, J., Vapnik, V.: Predicting Time Series with Support Vector Machines. In: Gerstner, W., Hasler, M., Germond, A., Nicoud, J.-D. (eds.) ICANN 1997. LNCS, vol. 1327, pp. 999–1004. Springer, Heidelberg (1997)
2. Tay, F.E.H., Cao, L.: Application of Support Vector Machines in Financial Time Series Forcasting. Omega 29, 309–317 (2001)
3. Zhang, J.F., Hu, S.S.: Chaotic Time Series Prediction Based on Multi-kernel Learning Support Vector Regression. ACTA Phys. Sinica (Chin. Phys.) 57, 2708–2713 (2008)

4. Guo, Y.M., Zhai, Z.J., Jiang, H.M.: Weighted Prediction of Multi-parameter Chaotic Time Series Using Least Squares Support Vector Regression (LS-SVR). J. Northwestern Polytechnical University 27, 83–86 (2009)
5. Zhang, X.R., Hu, L.Y., Wang, Z.S.: Multiple Kernel Support Vector Regression for Economic Forecasting. In: 2010 International Conference on Management Science and Engineering (17th), Melbourne, pp. 129–134 (2010)
6. Qi, M., Zhang, G.P.: Trend Time Series Modeling and Forecasting with Neural Networks. IEEE Trans. Neural Networks 19, 808–816 (2008)
7. Xi, J.H.: Study on Long-term Prediction Technology of Chaotic Time Series. Ph. D. Thesis of Dalian University of Technology (2005)
8. Vapnik, V.: Theory of Statistical Learning. In: Haykin, S. (ed.), pp. 375–563. A Wiley-Interscience Publication, John Wiley & Sons, Inc., New York (1998)
9. Suykens, J.A.K., Branbanter, J.K., Lukas, L., et al.: Weighted Least Squares Support Vector Machines: Robustness and Spare Approximation. Neurocomput. 48, 85–105 (2002)
10. Zhang, Y.: Enhanced Statistical Analysis of Nonlinear Process Using KPCA, KICA and SVM. Chem. Eng. Sci. 64, 801–811 (2009)
11. Xie, J.H.: LS-SVM Method Applied to Detect Damage for Piezoelectric Smart Structures. Sens. Actuat. 20, 164–167 (2007)
12. Xie, J.H.: Kernel optimization of LS-SVM Based on Damage Detection for Smart Structures. In: 2nd IEEE International Conference on Computer Science and Information Technology, Beijing, pp. 4244–4519 (2009)
13. Yang, L., Zhang, L., Zhang, S.X., Liu, J.W.: Comparison Research of Single Kernel and Multi-kernel Relevance Vector Machine. Comput. Engineer. 36, 195–197 (2010)
14. Remaki, L., Cheriet, M.: KCS-New Kernel Family with Compact Support in Scale Space: Formulation and Impact. IEEE Trans. Image Process. 9, 970–981 (2009)
15. Huang, X.: The Study on Kernels in Support Vector Machine. Master degree thesis of Suzhou University (2008)
16. Ye, M.Y., Wang, X.D., Zhang, H.R.: Chaotic Time Series Forecasting Using Online Least Squares Support Vector Machine Regression. ACTA Phys. Sinica 54, 2568–2573 (2005)