Predicting Noisy Data with an Improvement RBF Neural Network for Surrogate Models

Yaping Jiang[1, a], Guosong Wei[2, b], Xueyan Sun[2, c] and Yueqiang Zhang[1, d]

[1] China North Vehicle Research Institute, Beijing 100072, China;

[2] School of Mechanical Engineering, Dalian University of Technology, Dalian 116024, China.

[a]13911885540@163.com, [b]guos.wei@hotmail.com, [c] sunxueyan409@163.com, [d]syjwork@yeah.net

Keywords: Noise data, RBF, Neural network, Surrogate model.

Abstract. The process of noise data is a significant issue to the application of data. In this paper, we propose a method of processing noisy data based on improvement radial basis function (RBF) neural network to handle noise data. We establish surrogate models for two kinds of standard functions with noise data and noise-free data respectively, then by means of testing the two models with a set of perfect test data and analyzing the result of the comparative experiments to certify the effectiveness of this method.

1. Introduction

The processing of noisy data is one of the core issues that we can make use of data accurately and efficiently. Noisy data [1], also named meaningless data, is a sort of negative influence on original data because of the transfer mechanism of transferring data or external environment in the data transfer processing.

Currently, the existing methods for data process, such as binning, clustering and regression have some problems, including the complex of computing, heavy workload and the unsatisfied results. In an era of geometric increase of the amount of data, we need to develop some new methods to keep the pace of the progress. Radial basis function (RBF) neural network has been applied in almost every field because its strong ability to approximate nonlinear complex function and good fault tolerance .Zhang Xiaozai[2] applied RBF neural network in vessel radiated noise prediction, and the experiment proved that the effect is better than the traditional Volterra adaptive filter and also has a faster speed of convergence. Shen Yanxia[3] designed of sliding mode fault tolerant controller switch gain based on RBF neural network's ability of fault tolerance, and for wind energy conversion system fault diagnosis and fault-tolerant control online reorganization. Menita Carozza[4] presented an improved RBF neural network based on RBF which could be used to establish models even though the original data has been polluted within a certain range. DONG-MEI LI[5] applied RBF neural network in modeling of Chaotic systems with noise ,and its feasibility was verified by comparing approximate models between chaotic system and noisy chaotic system.

This paper is organized as follows: in the first part, we make a simple introduction about RBF neural network; then, we designed the experiments based on RBF neural network approximation model; in the third part, we analyze the results of experiments, and finally we give a conclusion.

2. RBF neural network

Radial basis function (RBF) neural network is a three layer forward neural network consisted of input layer, hidden layer and output layer. There are full connection between the hidden layer and the upper with lower level, and the transfer function of input layer to hidden layer called radial basis function, where, Gaussian function are selected as the radial basis function generally. Its structure is shown in figure 1.

After the establishment of the approximate model, we still need to test the approximate level on response of the approximate model to the true response. The principle of approximate model test is to

treat the observation value of design sample points as the true value, and treat the calculation value of RBF model as the estimated value, thus we compare the difference between the estimated value and the true value to judge the quality of the model. Testing method commonly uses approximate model including R-Squared, Adjusted R-Squared, RMSE, etc. Here we use R-Squared to make an error analysis and we will introduce it in next part.
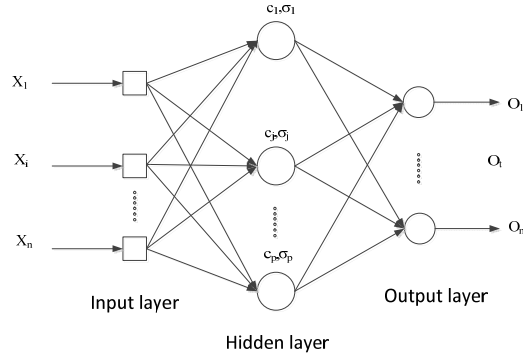

Fig. 1   RBF neural network architecture

3.   Design of experiment

In this paper, we manipulate radial basis function neural network to build surrogate model respectively for two kinds of ten-dimensional standard functions with noise-free data and data with noise, then employ a set of test data to test the two models, and provide the relative error by analyzing the experimental results. This section describes the specific design of experiments.

Sample point collection: In order to build the surrogate model, we usage the DOE matrix and Optimal-Latin-Hypercube method which could greatly improve the uniformity of Random Latin hypercube and increase accuracy of the fitting to generate N sample points. At the same time, manipulate optimization time-consuming.

Noise-free data generation: Selecting the sample points generated in DOE as the input of the two standard function to get noise-free data .Function expression as shown in Table 1.

Noisy data generation: Selecting m function values from noise-free data randomly as input of white noise function and setting SNR to generate noisy data.
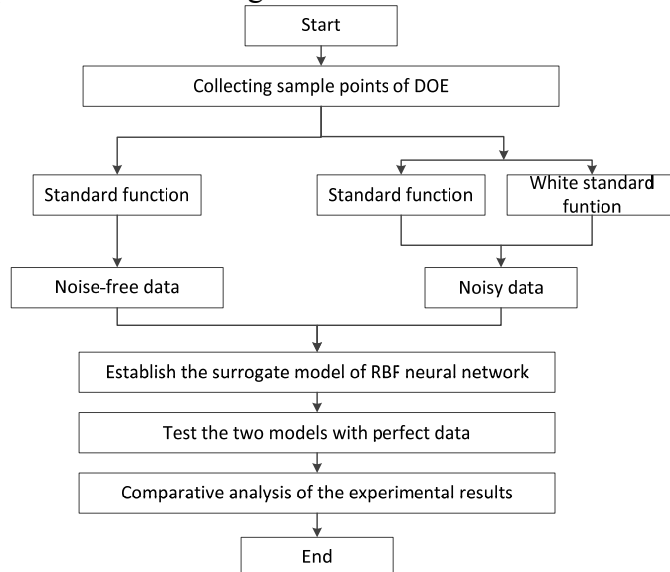
Experimental design process shown in Figure 2.

Fig. 2   Experimental design process

Table 1 Testing function		
Function	Function Formula	Range
1. Shifted Sphere	$f(x) = \sum\limits_{i=1}^{n} x_i^2$	[-10, 10]
2. Shifted Ellipsoid	$f(x) = \sum\limits_{i=1}^{n} i x_i^2$	[-10, 10]

4. Experimental results and analysis

We make experiment for function 1 (Shifted Sphere). Employing the DOE matrix and Optimal-Latin-Hypercube method, we get 200 sample points to generate the noise-free data and noisy data. Then we establish the approximate model based on RBF for the two sets of data respectively and evaluate approximate level with the R-Squared. The expression of R-Squared as follows:

$$R^2 = 1 - \frac{\sum\limits_{k=1}^{m}\left(\hat{y}(x^k) - \bar{y}(x^k)\right)^2}{\sum\limits_{k=1}^{m}\left(y(x^k) - \bar{y}(x^k)\right)^2}$$

（1）

Where, $y(x^k)$ is the actual response value of x^k, $\hat{y}(x^k)$ is the predict response value of x^k. It can be inferred by the formula that when the actual and predicted values are closer, $\dfrac{\sum\limits_{k=1}^{m}\left(\hat{y}(x^k) - \bar{y}(x^k)\right)^2}{\sum\limits_{k=1}^{m}\left(y(x^k) - \bar{y}(x^k)\right)^2}$ is smaller, and the value of R^2 is closer to 1, the value of R is closer to 1, which means the better performance of the surrogate model. Generally, we admit the approximate model is reliable when R>0.9.

Where, Data and Noisy-data are R-Squared of noise-free data and noisy data; according to the table 2, approximate models satisfy the conditions of credibility. Figures 3 and 4 are error analysis chart of its noise-free data and noisy data proxy models.

Table 2 R-Squared		
Function	Data	Noisy data
1	0.99586	0.98234
2	0.99219	0.9512

Thus, the two approximate models satisfy the conditions of credibility and could be used to predict.

We use Optimal-Latin-Hypercube method to generate 5 sets of testing data.Y1 is predictive value of perfect data model; Y2 is predictive value of noise data model. And R is the relative deviation by Y1.we consider it is feasibility when R<5%.
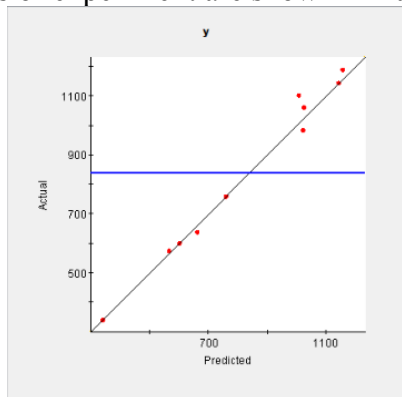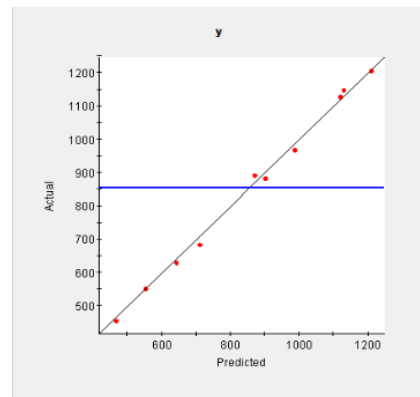
The results of experiment are shown in Table 3.

Fig.3　data error analysis

Fig. 4 Noisy data error analysis

Table3		Results of experiment of function 1			
Function1	P1	P2	P3	P4	P5
Y1	906.29	709.43	354.68	909.91	748.93
Y2	913.68	707.53	367.7	877.79	744.12
R	0.815%	0.296%	3.62%	3.53%	0.642%

The value of R is all less than 5%, it can be considered that the two models have the same approximate correspondence, which means that RBF neural network can establish approximate same agent model whether there exists noisy data or not within a certain range.

With the same method to test function 2, the results followed as Table 4:

Function 2 and function 1 show the same result that the value of are all less than 5% within the acceptable range.

Table4		Results of experiment of function 1			
Function2	P1	P2	P3	P4	P5
Y1	3392.6	2650.7	2894.4	2106.7	2281.0
Y2	3353.7	2678.1	2762.8	2097.4	2297.6
R	**1.14%**	1.03%	4.54%	0.441%	0.73%

5.   Conclusion

In this paper, we studied a noisy data processing method based on radial basis function (RBF) neural network. We build surrogate model of noise data with RBF neural network and apply this model in predicting data. Then we compared results with predict results of surrogate model of noise-free data. The result shows that the method of this paper provided is feasible for noise data processing.

References

[1]. Vanecek, P. Delaunay space division for RBF image reconstruction. Spring Conference on Computer Graphics. Budmerice, Slovakia, May, 2010, p 109-116.

[2]. Zhang Xiaozai and Yang Hongchao. To Predict the Radiation Noise of Ship Based on RBF Neural Network. Value Engineering. (2011) No. 14, p. 62-63.

[3]. Shen Yanxia, JI Lingyan, et al. Design of Fault-tolerant Controller for Wind Energy Conversion System Based on the RBF Neural Network fault observer. Information and Control. Vol. 4(2015)No.3, p. 359-366.

[4]. Carozza, M. and S. Rampone, Function approximation from noisy data by an incremental RBF network. Pattern Recognition, Vol. 32 (1999)No. 12: p. 2081-2083.

[5]. Li, D. and F. Li. Identification of chaotic systems with noisy data based on RBF neural networks. International Conference on Machine Learning and Cybernetics. Baoding, China, 2009, p. 2857-1859.