

The Quality Prediction in Small-batch Producing Based on Weighted Least Squares Support Vector Regression

Sheng Bo Zhang and Qing Ling Dai

Jiujiang University, Jiujiang, Jiangxi, 332005, China

Abstract. A novel quality prediction method with mobile time window is proposed for small-batch producing process based on weighted least squares support vector regression (LS-SVR). The design steps and learning algorithm are also addressed. In the method, weighted LS-SVR is taken as the intelligent kernel, with which the small-batch learning is solved well and the nearer sample is set a larger weight, while the farther is set the smaller weight in the history data. A typical machining process of cutting bearing outer race is carried out and the real measured data are used to contrast experiment. The experimental results demonstrate that the prediction accuracy of the weighted LS-SVR based model is only 20%-30% that of the standard LS-SVR based one in the same condition. It provides a better candidate for quality prediction of small-batch producing process.

1 Introduction

For small batch production, it is unrealistic to make trial runs for the determination of suitable values of the variables in machining operations [1,2]. This problem becomes even worse with the increasing requirements of product accuracy, the more frequent application of new materials and new machining operations, as, for instance, high-speed and precision machining of hardened steel components [3].

Many methods of quality prediction with the techniques are proposed such as liner predicting model, generalized liner predicting model, memory-based predictor [4] and neural networks [5,6]. Recently, support vector machines proposed by Vapnik have become a key machine learning technique [7]. They have strong theoretical foundation and show excellent empirical success in supervised learning framework [8]. Quality prediction can attribute to supervised learning framework. So support vector machines such as SVR, LS-SVR are applied to deal with quality prediction [9,10].

However, in actual production, quality parameters change constantly with process. The short-term impact, in future, on producing accuracy mainly comes from not history parameters but the latest processing parameters. We call the rule as “nearer data weigh heavier and farther data weigh lighter”. The concept of membership was introduced into the article, and a weight LS-SVR prediction model based time-domain membership was presented. This model put different memberships according to the impact extent of history data. It reduces the impact of early data to the current producing process, and improves the accuracy of real-time prediction. The model was researched in the producing process of

bearing’s outer ring, and compared prediction accuracy with the traditional model. The research demonstrates that the model is applicable and reasonable to small-batch producing process.

2 Principle of Wighted LS-SVR

2.1 Conventional LS-SVR

Consider a given training set of n data points $\{\mathbf{x}(i), y(i)\}_{i=1}^n$ with input $\mathbf{x}(i) \in R^m$ and output $y(i) \in R$. In feature space SVR models take the form as

$$y = \boldsymbol{\omega}^T \varphi(\mathbf{x}) + b, \quad (1)$$

where the nonlinear function $\varphi(\cdot)$ maps the input data into a higher dimensional feature space. Note that the dimension of $\boldsymbol{\omega}$ is not specified. The training is to estimate the optimal parameters $\boldsymbol{\omega}$, b in the regression function (1).

LS-SVR is a tool based on statistical learning theory and selects squared error ξ_i as the loss function [11]. Therefore, the following optimization problem is formulated as

$$\min_{\boldsymbol{\omega}, \xi} J(\boldsymbol{\omega}, \xi) = \min \frac{1}{2} \boldsymbol{\omega}^T \boldsymbol{\omega} + \frac{1}{2} \gamma \sum_{i=1}^n \xi_i^2, \quad (2)$$

subject to the equality constraints

$$y(i) = \boldsymbol{\omega}^T \varphi(\mathbf{x}(i)) + b + \xi_i, \quad i = 1, 2, \dots, n, \quad (3)$$

Zhang Shengbo : E-mail: B761127@tom.com

where γ is a positive constant parameter used to control the tradeoff between the training error and the model complexity. The larger γ is, the less error of regression model will make. This is the only free parameter in the LS-SVR formulation.

The parameter ω can be evaluated according to objective function (2) and constraints (3). This corresponds to a form of ridge regression and Lagrange function is given by

$$L(\omega, b, \xi, a) = \frac{1}{2} \omega^T \omega + \gamma \sum_{i=1}^n \xi_i^2 - \sum_{i=1}^n a_i (\omega^T \varphi(\mathbf{x}(i)) + b + \xi_i - y(i)), \quad (4)$$

with Lagrange multipliers a_i , $i = 1, \dots, n$.

The optimal $\mathbf{a} = [a_1, \dots, a_n]^T$ can be solved on Karush-Kuhn-Tucke conditions [2], which for optimality are

$$\begin{cases} \frac{\partial L}{\partial \omega} \rightarrow \omega = \sum_{i=1}^n a_i \varphi(\mathbf{x}(i)) \\ \frac{\partial L}{\partial b} \rightarrow \sum_{i=1}^n a_i = 0 \\ \frac{\partial L}{\partial \xi_i} \rightarrow a_i = \gamma \xi_i \\ \frac{\partial L}{\partial a_i} \rightarrow \omega^T \varphi(\mathbf{x}(i)) + b + \xi_i - y(i) = 0 \end{cases}. \quad (5)$$

After elimination of ξ_i , ω , the solution is solved by the following set of linear equations

$$\begin{bmatrix} 0 & \boldsymbol{\theta}^T \\ \boldsymbol{\theta} & \boldsymbol{\Omega} + \gamma^{-1} \mathbf{I} \end{bmatrix} \begin{bmatrix} b \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{Y} \end{bmatrix}, \quad (6)$$

where \mathbf{I} is the identity matrix, $\boldsymbol{\theta} = [1, \dots, 1]^T$, $\mathbf{a} = [a_1, \dots, a_n]^T$ and the Mercer condition

$$\boldsymbol{\Omega}_{ij} = \varphi(\mathbf{x}(i))^T \varphi(\mathbf{x}(j)) = \psi(\mathbf{x}(i), \mathbf{x}(j)), \quad (7)$$

has been applied.

Lagrange multiplier \mathbf{a} and model offset b could be solved in formula (6), which could also be expanded as

$$\begin{aligned} \boldsymbol{\theta}^T \mathbf{a} &= 0, \\ b \boldsymbol{\theta} + (\boldsymbol{\Omega} + \gamma^{-1} \mathbf{I}) \mathbf{a} &= \mathbf{Y}, \end{aligned}$$

namely,

$$\hat{\mathbf{a}} = (\boldsymbol{\Omega} + \gamma^{-1} \mathbf{I})^{-1} (\mathbf{Y} - b \boldsymbol{\theta}). \quad (8)$$

Substituting formula (8) into formula (6), and

$$\boldsymbol{\theta}^T \mathbf{a} = 0 \rightarrow \boldsymbol{\theta}^T (\boldsymbol{\Omega} + \gamma^{-1} \mathbf{I})^{-1} (\mathbf{Y} - b \boldsymbol{\theta}) = 0$$

$$\rightarrow \hat{b} = \frac{\boldsymbol{\theta}^T (\boldsymbol{\Omega} + \gamma^{-1} \mathbf{I})^{-1} \mathbf{Y}}{\boldsymbol{\theta}^T (\boldsymbol{\Omega} + \gamma^{-1} \mathbf{I})^{-1} \boldsymbol{\theta}}. \quad (9)$$

The finally results into the following LS-SVR model for function estimation

$$y = \sum_{i=1}^n \hat{a}_i \psi(\mathbf{x}, \mathbf{x}(i)) + \hat{b}. \quad (10)$$

2.2 Weighted LS-SVR

In this paper, we introduced the weight into standard LS-SVR. That is to introduce the weight terms w_i into each sample of LS-SVR to change the importance of input sample $\{\mathbf{x}(i), y(i), w_i\}_{i=1}^n$, where, $0 \leq w_i \leq 1$. Thus, the optimization objective function of Equation (2) is re-expressed as

$$\min_{\omega, \xi} J(\omega, \xi) = \min \frac{1}{2} \omega^T \omega + \frac{1}{2} \gamma \sum_{i=1}^n w_i \xi_i^2. \quad (11)$$

The Lagrange function is re-structured as

$$L(\omega, b, \xi, a, w) = \frac{1}{2} \omega^T \omega + \gamma \sum_{i=1}^n w_i \xi_i^2 - \sum_{i=1}^n a_i (\omega^T \varphi(\mathbf{x}(i)) + b + \xi_i - y(i)). \quad (12)$$

According to the optimum conditions, the matrix is also re-structured as

$$\begin{bmatrix} 0 & \boldsymbol{\theta}^T \\ \boldsymbol{\theta} & \boldsymbol{\Omega} + \gamma^{-1} \mathbf{W} \mathbf{I} \end{bmatrix} \begin{bmatrix} b \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{Y} \end{bmatrix}, \quad (13)$$

where $\mathbf{W} = [w_1^{-1}, w_2^{-1}, \dots, w_n^{-1}]$ is the weight vector.

Estimation function of weighted LS-SVR can be obtained by solving the weighted matrix (13) as

$$\begin{aligned} \hat{\mathbf{a}} &= (\boldsymbol{\Omega} + \gamma^{-1} \mathbf{W} \mathbf{I})^{-1} (\mathbf{Y} - b \boldsymbol{\theta}), \\ \hat{b} &= \frac{\boldsymbol{\theta}^T (\boldsymbol{\Omega} + \gamma^{-1} \mathbf{W} \mathbf{I})^{-1} \mathbf{Y}}{\boldsymbol{\theta}^T (\boldsymbol{\Omega} + \gamma^{-1} \mathbf{W} \mathbf{I})^{-1} \boldsymbol{\theta}}. \end{aligned} \quad (14)$$

3 Quality Prediction Using Weighted

LS-SVR

Let $x(t_0 + \tau)$, $x(t_0 + 2\tau)$, ..., $x(t_0 + n\tau)$ be a measured quality series with sampling time τ in a machining process. Our objective is to predict the future quality of the piecework from the series $x(t)$ according to the history quality data. However the quality series data are single-dimension usually, it is different to be used to the regression modeling directly.

For this problem, we apply the technique of mobile time window (MTW) of nonlinear quality series, the width of which is set to include $m+1$ data points [8]. In the model building, move the MTW on time series from the anterior to posterior, and the process of data extraction is shown as Figure 1.

The relationships in the MTW between $x(t_0 + k\tau)$ and $x(t_0 + (k-1)\tau)$ are supposed to be a nonlinear map

$$x(t_0 + k\tau) = f\{x(t_0 + (k-1)\tau), \dots, x(t_0 + (k-m)\tau)\} \quad (15)$$

where m is MTW width and also called embedding dimension. If m is small, the past values are not enough to predict the future value; however, if m is high, the training is hard.

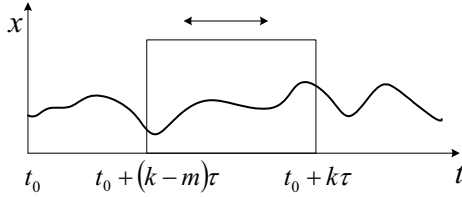


Figure 1. The process of MTW data extraction

Supposing that the current position of MTW is i , the input vector \mathbf{x}_i are the data from $i-1$ to $i-m$ on the quality series where \mathbf{x}_i denotes $x(t_0 + (i-1)\tau), \dots, x(t_0 + (i-m)\tau)$, and the output $y(i)$ is $x(t_0 + i\tau)$. When the MTW moving, the next training sample $\{\mathbf{x}(i+1), y(i+1)\}$ can be produced continually.

The training sample set are input into weighted LS-SVR to get the regression parameters \mathbf{a} and b . When the future value is predicted using the weighted LS-SVR-based model, the parameters \mathbf{a} and b are substituted in model (3) and the MTW is also shifted to the corresponding position. The history quality data $x(t_0 + (i-1)\tau), \dots, x(t_0 + (i-m)\tau)$ are taken as input data $\mathbf{x}(i)$ to obtain the model responses $\hat{y}(i)$, which are the prediction value of $x(t_0 + i\tau)$. The modeling process of quality series prediction with weighted LS-SVR is shown as Figure 2.

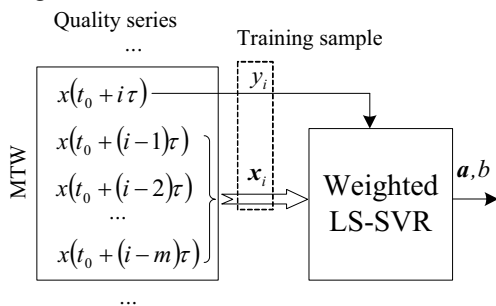


Figure 2. Quality series prediction modeling with wighted LS-SVR

The degree of importance and the impact on the future quality data by history data is increasing from far to near. Different membership to history samples according to its difference positions in the time domain

are observed. That is to say, nearer data samples are given larger weight, while further data samples are given smaller weight. Through this weight, in the quality series, the nearer samples are enhanced and the farther ones are weakened.

We apply the index distribution as the weight, and the actual weight of index distribution is available as follows

$$w_i = \alpha(1-\alpha)^{n-i}, \quad (16)$$

where α is index coefficient, $0 < \alpha < 1$ and $i=1,2,\dots,N$.

Form (16), the sum of weights is

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n w_i = \lim_{n \rightarrow \infty} \sum_{i=1}^n \alpha(1-\alpha)^{n-i} = 1. \quad (17)$$

The index coefficient α in time domain is very important to solve multivariate linear equation (15). It impacts on the accuracy of results obviously. Under normal circumstances, if the time sequence is stable, linear distribution and less value of α can be selected; if the data sequence fluctuates more seriously, index distribution and lesser value of α can be selected. The value of α can also be selected through optimized calculation with the least fitting error.

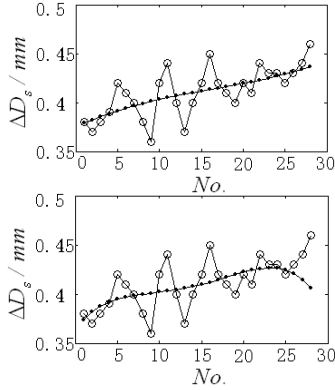
4 Experiments

An automatic turning process has been observed, and the real deviation sequence for outer ring of bearings size is shown in Table I. The tolerance of the outer ring is $90_{+0.30}^{+0.45}$ where, $\Delta D_{s \max s}$ is the largest deviation of the outer ring's diameter by measurement. In the turning process, there are system factors (such as tool wear, thermal deformation, etc.) and random factors (such as system vibration, rough material, etc.) impacting the turning quality.

Table 1. Real deviation sequence for outer ring size

| UNIT: MM | | | | | | | |
|-----------------------|------|------|------|------|------|------|------|
| No. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| $\Delta D_{s \max s}$ | 0.38 | 0.37 | 0.38 | 0.39 | 0.42 | 0.41 | 0.40 |
| $\Delta D_{s \min s}$ | 0.35 | 0.33 | 0.32 | 0.35 | 0.37 | 0.37 | 0.36 |
| No. | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| $\Delta D_{s \max s}$ | 0.38 | 0.36 | 0.42 | 0.44 | 0.40 | 0.37 | 0.40 |
| $\Delta D_{s \min s}$ | 0.35 | 0.32 | 0.39 | 0.40 | 0.35 | 0.33 | 0.34 |
| No. | 15 | 16 | 17 | 18 | 19 | 20 | 21 |
| $\Delta D_{s \max s}$ | 0.42 | 0.45 | 0.42 | 0.41 | 0.40 | 0.42 | 0.41 |
| $\Delta D_{s \min s}$ | 0.38 | 0.41 | 0.38 | 0.36 | 0.38 | 0.39 | 0.36 |
| No. | 22 | 23 | 24 | 25 | 26 | 27 | 28 |
| $\Delta D_{s \max s}$ | 0.44 | 0.43 | 0.43 | 0.42 | 0.43 | 0.44 | 0.46 |
| $\Delta D_{s \min s}$ | 0.38 | 0.37 | 0.40 | 0.38 | 0.36 | 0.40 | 0.41 |

Firstly, different prediction models using three-order polynomial regression, four-order polynomial regression, BP neural network, standard LS-SVR, and weighted LS-SVR were set up respectively, with the first twenty-five data in Table 1. Then the values of the 26th, 27th, 28th work pieces were predicted and the results were compared with the real ones.



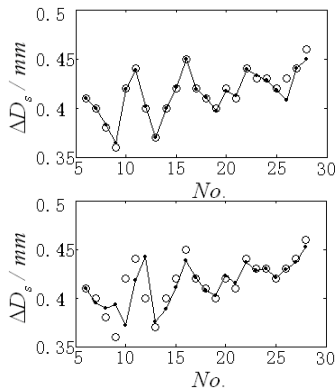
(a) 3-order polynomial based model; (b) 4-order polynomial based model;

Figure 3. Prediction using polynomial based model

In the experiments, the width of time was all set as 5. The polynomial and BP neural network are realized with the polynomial function and artificial neural network toolbox in Matlab6.5, respectively. In the BP neural network based model, the learning rate $a = 0.1$, hidden layer neurons were taken 6 and training steps is 10000. The prediction results using polynomial based model are shown in Figure 3.

In the testing, the data from 3th to 25th are taken as regression checking, and the data from 26th to 28th are taken as the testing. From Figure 3, it is obvious to see that polynomial model can predict quality deviation caused by system factors but cannot predict those by random factors in turning process.

Standard LS-SVR and weighted LS-SVR algorithm were achieved by matrix function in matlab6.5, with the parameter adjusting constant valued 500. RBF was chosen as Kernel function of above two models with the parameter $\sigma = 0.1$, and in the weighted LS-SVR based model, the distribution parameters $\alpha = 0.3$. The predicted deviations of LS-SVR and weighted LS-SVR based model are shown in Figure 4.



(a) Standard LS-SVR based model; (b) Weighted LS-SVR based model;

Figure 4. Prediction results using different LS-SVR based model

As seen from Figure 4, LS-SVR model has high regression precision for training samples. Output sequence of the LS-SVR model almost coincides with the real deviation sequence. There are more regression errors on the first half of sample sequence based on weighted LS-SVR model but much higher precision on the second half of the sample sequence. It's better to predict the three future data by weighted LS-SVR based model.

$$\text{Mean square error } MSE = \sum_{i=1}^n (\hat{y}_i - y_i)^2 / n \text{ was}$$

introduced to compare the prediction precision of each model, where \hat{y}_i , y_i and n are output of models, output of training sample sets, and the number of training samples, respectively. The prediction precisions of each model are shown in Table II.

Table 2. Prediction precision of different model

| Method | | 26th | 27th | 28th | MSE |
|--------------------|------------|-------|-------|-------|-------|
| Real value | | 0.43 | 0.44 | 0.46 | / |
| 3-order polynomial | Prediction | 0.431 | 0.434 | 0.437 | 0.010 |
| | Error | 0.001 | 0.006 | 0.023 | / |
| 4-order polynomial | Prediction | 0.421 | 0.415 | 0.406 | / |
| | Error | 0.009 | 0.025 | 0.054 | 0.029 |
| Standard LS-SVR | Prediction | 0.408 | 0.441 | 0.450 | / |
| | Error | 0.022 | 0.001 | 0.010 | 0.011 |
| Weighted LS-SVR | Prediction | 0.429 | 0.436 | 0.452 | / |
| | Error | 0.001 | 0.004 | 0.008 | 0.004 |

Analyzing the prediction results shown in Table III, we can contend that in all these experiments, weighted LS-SVR based model outperforms its main competitors – polynomial and standard LS-SVR based models. The MSE of weighted LS-SVR prediction model is only 20%-30% that of standard ones. That proves the rule: “nearer data weigh heavier and farther data weigh lighter”.

5 Conclusions

The main benefits of this presented prediction model are that it could be applied to quality prediction of small-batch producing process well and make history data ‘nearer is more weight’ through using the index distribution as the weight in time domain.

The testing result of weighted LS-SVR based prediction model is very competitive in comparison with the other models. Thus the presented prediction model is demonstrated to be capable of describing the variety of quality with time.

References

1. C.A. Van Luttervelt, and J. Peng, “Symbiosis of modelling and sensing to improve the accuracy of workpieces in small batch machining operations,”

- International Journal of Advanced Manufacturing Technology, 1999, **vol. 15**, no. 10, pp. 699-710.
2. D.M. Hawkins, and K.D. Zamba, "On Small Shifts in Quality Control," *Quality Engineering*, Sep. 2003, **vol. 16**, no. 1, pp. 143-149.
 3. H.A. Kishawy, "An experimental evaluation of cutting temperatures during high speed machining of hardened D2 tool steel", *Machining Science and Technology*, 2002, **vol. 6**, no. 1, pp. 67-79.
 4. E.S. Chng, S. Chen, and B. Mulgrew, "Gradient radial basis function networks for nonlinear and nonstationary time series prediction," *IEEE Transactions on Neural Networks*, **vol. 7**, no.1, pp. 190-194, 1996.
 5. Y. Becerikli, and Y. Oysal, "Modeling and prediction with a class of time delay dynamic neural network," *Applied Soft Computing Journal*, Aug. 2007, **vol. 7**, no. 4, pp. 1164-1169.
 6. E. Gomez-Ramirez, and K. Najim, E. Ikonen, "Forecasting time series with a new architecture for polynomial artificial neural network," *Applied Soft Computing Journal*, Aug. 2007, **vol. 7**, no. 4, pp. 1209-1216.
 7. V.N. Vapnik, *The nature of statistical learning theory*. New York: Springer-Verlag, 1999.
 8. V.N. Vapnik, "An overview of statistical learning theory," *IEEE Transaction Neural Networks*, 1999, **vol. 10**, no.5, pp. 988-999.
 9. J. A. K. Suykens, and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, 1999, **vol. 9**, no. 3, pp. 293-300.
 10. J. A. K. Suykens, J. Vandewalle, "Sparse least squares support vector machine classifiers," *European Symposium on Artificial Neural Networks*, 2000, Bruges Belgium, 37-42.
 11. J.A.K. Suykens, L. Lukas, and J. Vandewalle, "Sparse approximation using least squares support vector machines," *Proceedings - IEEE International Symposium on Circuits and Systems*, Geneva, **vol. 2**, pp. 757-760, 2000.