ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 109-114

Time Series Analysis for Quality Improvement: a Soft Computing Approach

K. Xu and S.H. Ng

Department of Industrial and Systems Engineering
National University of Singapore, Singapore 117576.

S.L. Ho

The Centre for Quality and Innovation, Ngee Ann Polytechnic, Singapore, 599489

Abstract

Quality improvement provides organizations with significant opportunities to reduce costs, increase sales, provide on time deliveries and foster better customer relationships. The design and manufacturing are among the critical processes for continuous quality improvement. Time series data collected from these processes are the useful source. While there are various techniques to explore these processes, Neural Networks (NN) approach is deemed as a promising alternative. However, as NN is a relatively new approach in quality engineering which is traditionally dominated by statistical analysis, there is still much doubt in its effectiveness compared with statistical modeling. The main focus here then is to construct a statistically reliable neural network model with an appropriate architecture to conduct the time series analysis. The purpose of this paper is thus two-fold. Firstly we develop the statistical interval analysis for neural network models which provide a statistical guide towards a reliable modeling architecture. Secondly, we apply the developed approach for quality improvement in various industries.

Key words: Quality Improvement, Time Series Analysis, Neural Networks, Soft Computing.

1. Introduction

Quality improvement provides organizations with significant opportunities to reduce costs, increase sales, provide on time deliveries and foster better customer relationships. Such improvements can contribute considerably to the bottom line of these organizations. The design and manufacturing are among the critical processes for continuous quality improvement. While there are various qualitative techniques to explore these processes, the most common approach is a quantitative approach. As such, the critical quality data collected in these processes are primary sources for further investigations into potential causes of poor quality. Furthermore, these data can provide crucial information for decision making. Interestingly, many of these

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 109-114

quality data are time series data. For example, automobile engine manufacturers are interested in the failure patterns of its engines over time as it provides important information for the life time predictions, warranty costs evaluations, and maintenance scheduling of its engines. With the rapid advances in microelectronics packaging technology, electronics assembly manufacturers are collecting crucial on-line process monitoring data to dynamically track quality measures. In semiconductor and materials industries, modeling the experimental data set of chemical vapor deposition process offers an effective approach for the off-line process quality improvement.

In various nonlinear modeling for time series analysis approaches, neural networks (NN) approach is deemed as a promising alternative. There are many successful applications that are widely reported. However, as NN is a relatively new approach in quality engineering, there is still much doubt in its effectiveness compared with traditional statistical modeling. The main focus here then is to construct a statistically reliable neural network model with an appropriate architecture to conduct the time series analysis. The purpose of this paper is thus two-fold. Firstly we develop the statistical interval analysis for neural network models which provide a statistical guide towards a reliable modeling architecture, particularly when only a small number of training samples are available. Secondly, we apply the developed approach for quality improvement in various industries.

The organization of this paper is as follows. Section 2 introduces a general framework for the time series modeling approach using neural networks. Section 3 develops the statistical interval analysis for NN models. Using the proposed approach for quality improvement, an industry application is presented in section 4. The conclusions are presented in section 5.

2. General framework for time series modeling of neural networks

A general time series forecasting model can be formulated as

$$X_{t+1} = f\,(X'_{t-m}, A, B)$$

where $\{X'_{t-m}\,;\ m = 0, 1, 2, \ldots, p\}$ represents a time series of lagged variables; A denotes the external explanatory variables, i.e. variables on which the series is thought to have a dependence if any; and B represents the random error variables. X_t can be represented in different forms. As the modeling progresses and once the NN model has adequately understand the underlying characteristics of the time series, future outcomes can be predicted when presented with new data patterns.

3. Statistical interval analysis for neural networks

computing the statistical intervals such as the prediction intervals allow us to quantify these uncertainties and therefore provide more information. In prediction tasks, it is desirable to construct confidence bounds upon the resulting point forecasts.Our approach is generally following the line of the non-liner regression. However, we improve the traditional error variance estimation method which may result in errors when only limited samples are available. Furthermore, we propose to use the

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 109-114

statistical interval analysis and coverage assessment as a guide to select the appropriate NN architecture. Hence, the constructed model may be statistical reliable.

3.1. Quantifying Uncertainties via Non-linear Regression

Given a general multi-response nonlinear model:

$$\mathbf{y} = \mathbf{f}\left(\mathbf{x};\boldsymbol{\theta}^*\right) + \boldsymbol{\varepsilon},$$

the least-squares estimate of true value $\boldsymbol{\theta}^0$ of an actual process is $\hat{\boldsymbol{\theta}}$, which is obtained by Levenberg-Marquardt (LM) algorithm through minimizing the error function (1) in Bayesian regularization. The error term $\boldsymbol{\varepsilon}$, is assumed to be $\boldsymbol{\varepsilon}_D$: prediction error vector from training data sets.

Given the prediction from the model as $\hat{\mathbf{y}} = \mathbf{f}\left(\mathbf{x};\hat{\boldsymbol{\theta}}\right)$, $\mathbf{f}\left(\mathbf{x};\hat{\boldsymbol{\theta}}\right)$ can be approximated in terms of $\mathbf{f}\left(\mathbf{x};\boldsymbol{\theta}^*\right)$ by first-order Taylor expansion:
$\hat{\mathbf{y}} = \mathbf{f}\left(\mathbf{x};\hat{\boldsymbol{\theta}}\right) \approx \mathbf{f}\left(\mathbf{x};\boldsymbol{\theta}^*\right) + [\nabla\mathbf{f}\left(\mathbf{x};\boldsymbol{\theta}^*\right)]^T\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right)$

where

$$\nabla\mathbf{f}\left(\mathbf{x};\boldsymbol{\theta}^*\right) = [\frac{\partial}{\partial\theta_1}\mathbf{f}\left(\mathbf{x};\boldsymbol{\theta}^*\right) \quad \frac{\partial}{\partial\theta_2}\mathbf{f}\left(\mathbf{x};\boldsymbol{\theta}^*\right) \quad ... \quad \frac{\partial}{\partial\theta_p}\mathbf{f}\left(\mathbf{x};\boldsymbol{\theta}^*\right)]^T$$

where *p* is the number of parameters.

The difference between the true value $\mathbf{y}^0$ and the predicted value $\hat{\mathbf{y}}^0$ for points that are not used to train the model is given by:

$$\begin{aligned}\mathbf{y}^0 - \hat{\mathbf{y}}^0 &\approx \mathbf{y}^0 - \mathbf{f}\left(\mathbf{x}_0;\boldsymbol{\theta}^*\right) - [\nabla\mathbf{f}\left(\mathbf{x}_0;\boldsymbol{\theta}^*\right)]^T\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right)\\ &= \boldsymbol{\varepsilon}_D^0 - [\nabla\mathbf{f}\left(\mathbf{x}_0;\boldsymbol{\theta}^*\right)]^T\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right)\end{aligned}$$

The expectation of the difference is: $E\left[\mathbf{y}^0 - \hat{\mathbf{y}}^0\right] \approx E\left(\boldsymbol{\varepsilon}_D^0\right) - [\nabla\mathbf{f}\left(\mathbf{x}_0;\boldsymbol{\theta}^*\right)]^T E\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right) = \mathbf{0}$

Assuming statistical independence between $\boldsymbol{\varepsilon}$ and $\hat{\boldsymbol{\theta}}$, the variance can be expressed as:

$$\begin{aligned}\mathrm{Var}\left[\mathbf{y}^0 - \hat{\mathbf{y}}^0\right] &\approx \mathrm{Var}[\boldsymbol{\varepsilon}_D^0] + \mathrm{Var}\left\{[\nabla\mathbf{f}\left(\mathbf{x}_0;\boldsymbol{\theta}^*\right)]^T\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right)\right\}\\ &\approx \sigma_D^2\mathbf{I} + \mathrm{Var}\left\{[\nabla\mathbf{f}\left(\mathbf{x}_0;\boldsymbol{\theta}^*\right)]^T\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right)\right\}\\ &\approx \sigma_D^2\mathbf{I} + \nabla\mathbf{f}\left(\mathbf{x}_0;\boldsymbol{\theta}^*\right)^T\mathrm{Var}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right)\nabla\mathbf{f}\left(\mathbf{x}_0;\boldsymbol{\theta}^*\right)\end{aligned}$$

The estimates for the parameters of NN model obtained via Levenberg-Marquardt (LM) algorithms are:

$$\hat{\boldsymbol{\theta}} = \mathbf{H}^{-1}\mathbf{J}^T\left(\boldsymbol{\varepsilon}' + \mathbf{J}\boldsymbol{\theta}^*\right),$$

where, $\mathbf{H}$ is the Hessian matrix and $\mathbf{J}$ is the Jacobian matrix of the training sets error.

Therefore, the asymptotic variance of the estimated parameters can be approximated by:

$$\mathrm{Var}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right) \approx s^2\mathbf{H}^{-1}\mathbf{J}^T\mathbf{J}\mathbf{H}^{-1}$$

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 109-114

For large samples, we have approximately (Seber and Wild, 1989):

$$\frac{\mathbf{y}^0 - \hat{\mathbf{y}}^0}{\sqrt{\mathrm{Var}\left[\mathbf{y}^0 - \hat{\mathbf{y}}^0\right]}} \sim t_{n-\delta}$$

where, $t_{n-\delta}$ is student *t*-distribution with $n-\delta$ degrees of freedom; *n* is the number of training. We will discuss the selection of δ in following section.

It follows that an approximate $100(1-\gamma)$% prediction interval for the predicted value $\hat{\mathbf{y}}^0$ is given by:

$$\hat{\mathbf{y}}^{\mathbf{0}} \pm t_{n-\delta}^{\gamma/2}\left(\sqrt{\sigma_D^2\mathbf{I} + s^2[\nabla\mathbf{f}\left(\mathbf{x}_0;\boldsymbol{\theta}^*\right)]^T\mathbf{H}^{-1}\mathbf{J}^T\mathbf{J}\mathbf{H}^{-1}[\nabla\mathbf{f}\left(\mathbf{x}_0;\boldsymbol{\theta}^*\right)]}\right)$$

3.2. Error Variance Estimation

In the discussion of De Veaux *et al*. (1998), for conventional MLP and NN with weight decay method, the estimated error variance is given by $s^2 = \frac{E_D}{n-p}$, where p is the number of parameters which could be very large and, in some cases, may be larger than *n;* such that estimation of prediction interval is impossible. By stopping the algorithm before convergence, the weights or bias are shrunk toward 0, thus reducing the effective number of parameters in the models. However, this would result in overly conservative prediction intervals.

NN model using Bayesian Regularization (BR) is broadly used to keep the over-fitting problem in check (MacKay (1992) and Foresee and Hagan (1997)). Thus, our approach in using the effective parameter δ estimated from the BR process is a reasonable alternative to the statistical interval analysis. In MacKay (1992) and Foresee and Hagan (1997), $\delta = p - 2\alpha\mathrm{Trace}(\mathbf{H})^{-1}$ is the defined as the effective number of parameters; *p* is the number of parameters of NN. **H** is the Hessian matrix of the objective function $F(\theta) = \beta E_D + \alpha E_\theta$ where $E_D = \left(\mathbf{y} - \mathbf{y}^*\right)^T\left(\mathbf{y} - \mathbf{y}^*\right)$ is the residual sum of squares, $E_\theta = \boldsymbol{\theta}^T\boldsymbol{\theta}$ is the sum of squares of the network coefficients (weights and bias), and α and β are the objective function parameters whose relative size determines the training direction. Therefore, in our statistical interval formulation, $\sigma_D^2 \approx s^2 = \frac{E_D}{n-\delta}$. It implies that the actually number of weights and bias (effective parameters in the models) would be significantly smaller when using the NN modeling with BR. Therefore, it may effectively prevent over-fitting of models when sparse data sets are available for training.

3.3. Statistical Guide of Selecting the NN Architecture Parameters

The probability of coverage of the prediction interval is approximately equal to the nominal $(1-\gamma)$ with large sample sizes (asymptotic inference). This can be evaluated

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 109-114

empirically using Monte Carlo simulation. We conduct 5000 simulations using different NN models trained with extra random error to response data settings. The coverage percentage is calculated based on all run designs. The random noise is assumed to be normal with mean 0 and a small standard deviation (in practice, values of σ near 0.1 are common (De Veaux *et al*., 1998)). Taking into account the noise in the training data set, a more suitable measure would be $\sigma = 0.1 + 1/\beta$, where β is the noise of the training data sets and readily available from the BR process. The coverage probability can then be estimated from the simulation as : Coverage probability $= \frac{\text{number of predicted points within intervals}}{5000 * (\text{number of experiment run})}$. Therefore, we can use the statistical interval as well as the coverage probability as the measure to select the appropriate the NN parameters such as the number of the hidden units etc. This way, the model constructed is statistical reliability in some sense.

4. Cases Study: Process modeling and improvement

In semiconductor and material industries, chemical vapor deposition (CVD) process is the essential technique for depositing silicon nitride, silicon dioxide, polysilicon, refractory metals etc.(Jaeger, 1988) or materials synthesize . Plasma-enhanced CVD (PECVD) improves CVD by allowing the process to form a thin film on a substrate surface at the relatively low temperature. Fig. 1 shows a typical layout of a plasma CVD apparatus with a parallel plate electrode structure. A substrate is placed on the grounded electrode. The reaction gas is supplied from the opposite plate in order to form a uniform film.

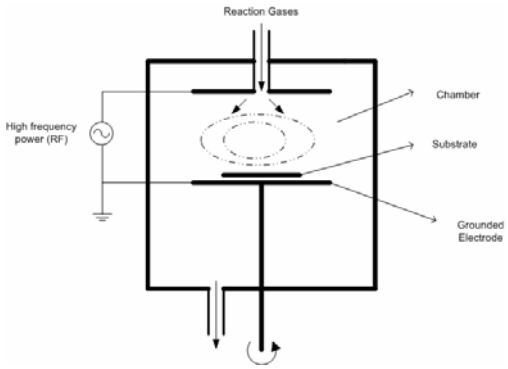

Figure 1. The typical layout of a plasma CVD apparatus

As a widely used technique, modeling and optimizing this process has received much attention for many years. In this example, Eight factors: x_1: cleaning method; x_2: chamber temperature; x_3: batch after cleaned chamber; x_4: flow rate of SiH_4; x_5: flow rate of N_2; x_6: chamber pressure; x_7: R.F.power; x_8: deposition time are selected to analyze two quality characteristics: y_1, the deposition thickness and a refractive index y_2. The total number of the training samples is 16. Due to small number of

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 109-114

samples, a statistical interval analysis should be used to provide a reliable modeling. Here, we select the three layers neural networks of BR with 8 input nodes, 2 output nodes and 12 hidden layers based on the proposed prediction interval analysis and coverage assessment process. The sum of squares errors for two responses are 2.6586e+004 (y_1) and 0.0660 (y_2). The coverage is 94.8% for y_1 and 98.8% for y_2. The average half the length of 95% prediction intervals is 326.104 for y_1 and 0.514 for y_2. Note that final solutions are very close to the defined targets of the two responses (y1=1000, y2=2.0). Note that the developed neural network method can be easily adjusted to modeling different controls factors and quality measures of processes.

Table 1. Best settings for PECVP Experiments

x_1	x_2	X_3	x_4	x_5	x_6	x_7	x_8	Y_1	y_2
2.0	1.17	1.0	1.0	1.0	1.0	3.0	3.0	995.48	2.0090
2.0	1.18	1.0	1.0	1.0	1.0	3.0	3.0	994.71	2.0083

5. Conclusions

This paper investigates the application of time series approach under the umbrella of soft computing, particularly neural networks, to modeling and improve the process quality. A statistical interval analysis approach is developed and thus the appropriate neural networks architecture can be selected to provide reliable modeling for quality improvement.

References

1. De Veaux, R. D., Schumi, J., Schweinsberg, J. and Unger, L. H.,. "Prediction Intervals for Neural Networks via Nonlinear Regression", *Technometrics*, vol. 40 (4), 273-282. (1998)

2. Foresee, F.D. and Hagan, M.T.. "Gauss-Newton Approximation to Bayesian Learning". in *IEEE International Conference on Neural Networks*, Houston, June, USA, 1930-1935. (1997)

3. Hwang, J. T. G. and Ding, A. A.. "Prediction Intervals for Artificial Neural Networks", *Journal of the American Statistical Association*, 92, 748-757. (1997)

4. Jaeger, R. *Introduction to Microelectronic Fabricatio*n. Addison-Wesley, New York. (1988)

5. MacKay, D.J.C.. "Bayesian Interpolation". *Neural Computation* 4, 415-447. (1992)

6 Seber, G.A.F. and Wild, C.J., *Nonlinear Regression* , John Wiley & Sons, New York. (1989)