Neural network time series classification of changes in nuclear power plant processes

Karel Kupka[1], Milan Meloun[2]
[1]TriloByte Statistical Software, Research Center for Quality and Reliability, St. Hradiste 300, 53352 Pardubice, Czech Republic, kupka@trilobyte.cz
[2]University Pardubice, Faculty of Chemistry, Analytical Chemistry Dept., Pardubice, Czech Republic

Abstract

Time series are typical data output from technological processes. Diagnostics of process data such as model change detection, outlier detection, etc. [1] are often of primary interest for quality management. For autocorrelated processes, many models and procedures have been suggested, many of them based on uni- and multivariate EWMA, AR, ARIMA, CUSUM. Two types of models for stationary univariate series were tested: linear partial least squares autoregresison (PLSAR) and nonlinear perceptron-type feed-forward neural network autoregression (ANNAR) with multistep prediction [5, 7, 8]. Here we describe mainly the ANN based models [2]. Lack of knowledge of statistical properties of prediction is solved by multiple overlay estimates, which makes it possible to assess and predict heteroscedastic variance and construct statistical models and control charts of processes with confidence intervals. It is shown that orthogonal PLSAR models are more stable than ANNAR and its parameters could be used to identify and classify different physical modes of processes. However, though neural network models often behave as unstable black boxes, our results suggest that the ANN solutions form nonconvex subspaces of a limited dimensionality that can be used to classify different types of time series using support vector machines (SVM) classifiers. Expansions of the proposed models for multivariate and non-stationary data are also possible and allow to control and classify more complex processes in technology. Throughout this study, the commercial statistical software QCExpert 3.2 was used [13].

Key words: time series, neural network, classification, support vector machines, prediction, QCExpert

1. Introduction

Better understanding technological processes may provide useful information needed to improve economy and quality in industry. Extensive process studies were performed on dynamic processes in a sub-critical water based primary circuit of a nuclear power plant near Brno, Czech Republic, see Fig. 1. Process parameters data including temperature, pressure, flowrates, conductivity were collected in form of time series as illustrated on Fig. 2. These processes are highly turbulent and could not be well described by simple low order AR or ARIMA models. Moreover, the processes could switch between semi-stable physical states represented by different models. Better results were obtained using nonlinear multilayer neural network k-th degree autoregression with sigmoidal activation function,

$$x_n = \sigma_2 \left\{ \sum_{i=1}^{l} \alpha_i \sigma_{1i} \left(\sum_{j=1}^{k} \beta_{ij} x_{n-j} \right) \right\} \tag{1}$$

where x are the standardized time series data or their first or second differences to elliminate trends, α, β are linear coefficients (weights) and σ is the activation function,

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{2}$$

Least squares objective function (3) and Gauss-Newton optimization was used to train the network (optimize parameters) using the series data, different set of parameters was assigned to each series. In situations where there are suspected outliers in the series or the distribution is expected to be highly non-normal, least squares may lead to biased prediction and $1 < r < 2$ is recomended to be used instead of least squares with $r = 2$ to robustify the model.

$$G(\mathbf{a},\mathbf{b}) = \sum_{n=k+1}^{m} \left\{ x_n - \sigma_2 \left[\sum_{i=1}^{l} a_i \sigma_{1i} \left(\sum_{j=1}^{k} b_{ij} x_{n-j} \right) \right] \right\}^r \tag{3}$$

where m is the series length, **a**, **b** are estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ respectively and r is the exponent, for least squares, $r = 2$. Number of the parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, namely $p = l + lk$ is the dimension model parameter space.

Figure 1: The power plant, measurement site and the measuring gauges arrangement

2. ANN Time Series Modelling

It is well known that optimized ANN parameters tend to be very unstable and their values depend e.g. on the optimization procedure tuning and starting parameter values [9]. It was expected however, that the parameters will behave similarly to the parameters of poorly conditioned (e.g. multicollinear) linear regression model, where they lie in a subspace of the parametric space (rather then in a point). To verify this idea expressed by (4), the parameters α, β were computed many times (typically 200 - 500) beginning with different random starting values each time for a chosen series. Each time a totally different neural network was obtained as shown on Fig. 3. Principal component analysis was then applied to the set of these bootstrapped parameters. As shown on Fig. 4, significant drop in the scree plot and the dominating directions in the PCA-biplot strongly suggest that the dimensionality of solution space is substantially lower than that of the original model parameter space [6].

$$\begin{pmatrix} \dfrac{\partial \alpha_1}{\partial \alpha_1} & \cdots & \dfrac{\partial \alpha_1}{\partial \alpha_k} \\ \vdots & \dfrac{\partial \alpha_i}{\partial \alpha_j} & \vdots \\ \dfrac{\partial \alpha_k}{\partial \alpha_1} & \cdots & \dfrac{\partial \alpha_k}{\partial \alpha_k} \end{pmatrix} \text{subject to } \frac{\partial R(\alpha)}{\partial \alpha_i} \leq \delta; \delta \geq 0 \quad (4)$$

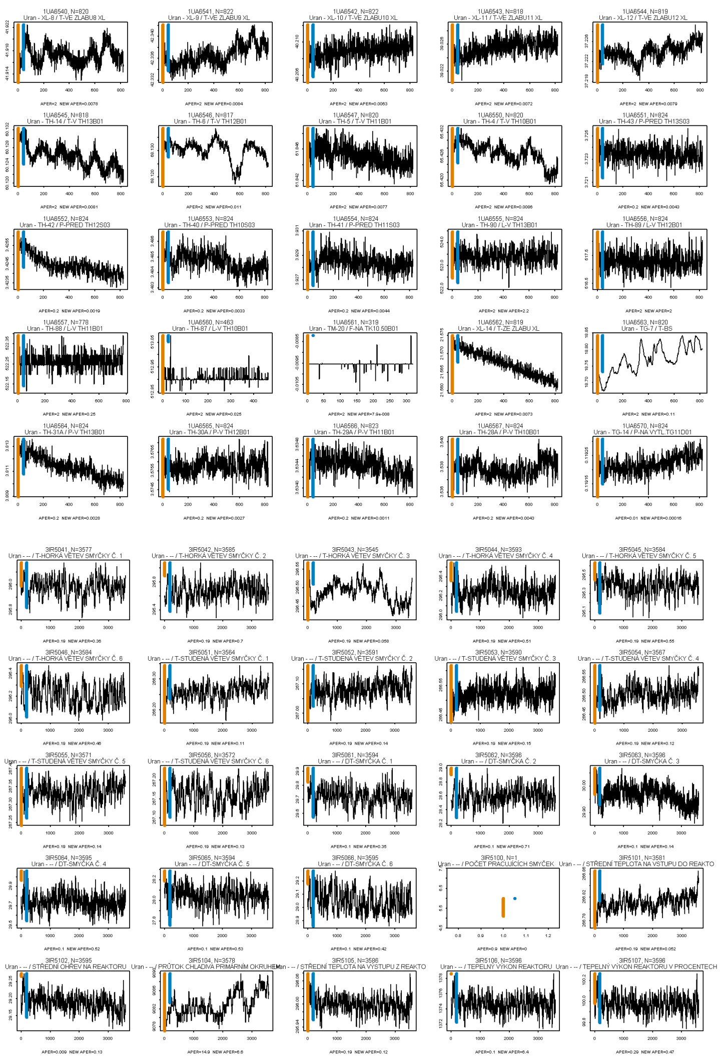

Figure 2: Different time series data under analysis

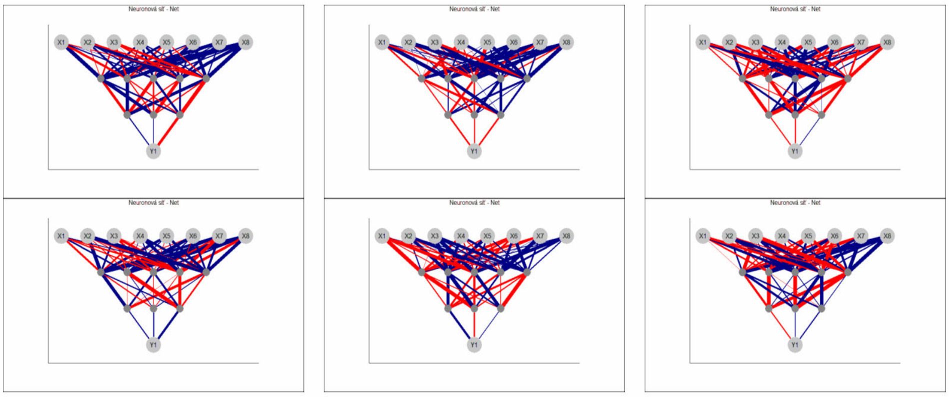

Figure 3: Six optimized nets with very different parameters fitted to the same series with practically the same residual squares sum. Color of the radii correspond to the sign of a weight, thickness is proportional to its absolute value.

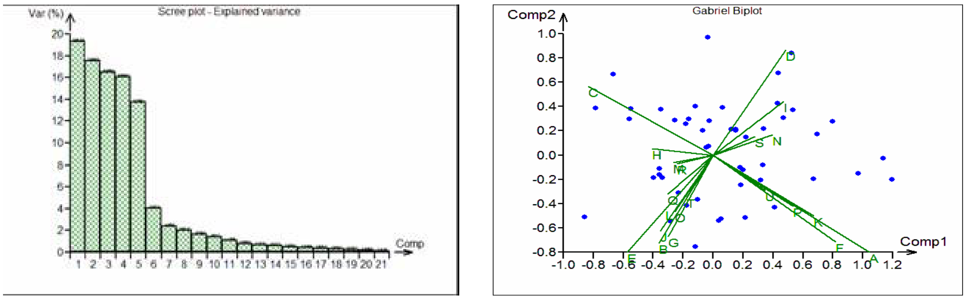

Figure 4: Principal component analysis (scree plot and PCA Biplot) reveal substantially lower dimensionality of the solution.

3. ANN Model Parameters as Classifiers

Since It is suggested that the first significant principal components be used as classifiers. These components occupy a (possibly non-convex) subspace of a dimension $p^* < p$ in an original parameter space as illustrated on Fig. 5 and their orthogonality may provide good predictive properties. In this study, we used an SVM (Support Vector Machine) classification model with RBF (Radial Base Function) kernel (5) [3, 4, 8, 10, 11, 12] . Less then 10, typically 5, principal components were satisfactory to discriminate between different time series data from distinct technological situations. Even in the first two principal component from the misclassification rate was under 20%. For 5 principal components the misclassification rate was under 2% as illustrated on Table 1.

$$\begin{aligned} \min_{\mathbf{w},b,\boldsymbol{\xi},\rho} \quad & \frac{1}{2}\mathbf{w}^T\mathbf{w} - \nu\rho + \frac{1}{l}\sum_{i=1}^{l}\xi_i \\ \text{subject to} \quad & y_i(\mathbf{w}^T\phi(\mathbf{x}_i)+b) \geq \rho - \xi_i, \\ & \xi_i \geq 0, i = 1,\ldots,l, \rho \geq 0. \end{aligned} \tag{5}$$

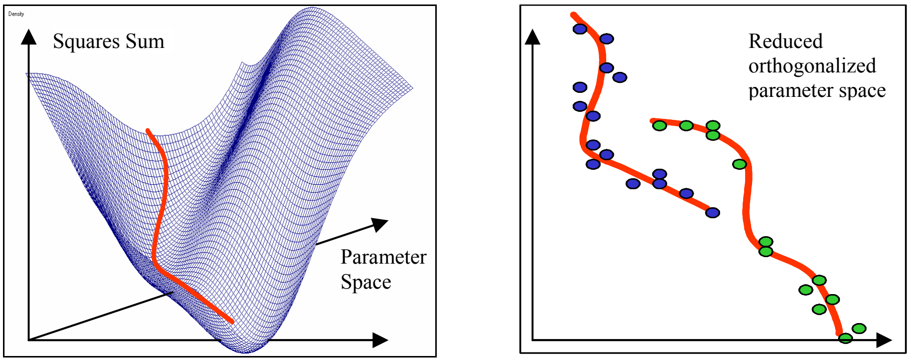

Figure 5: Subspace (red) of optimal ANN solutions of a time series data

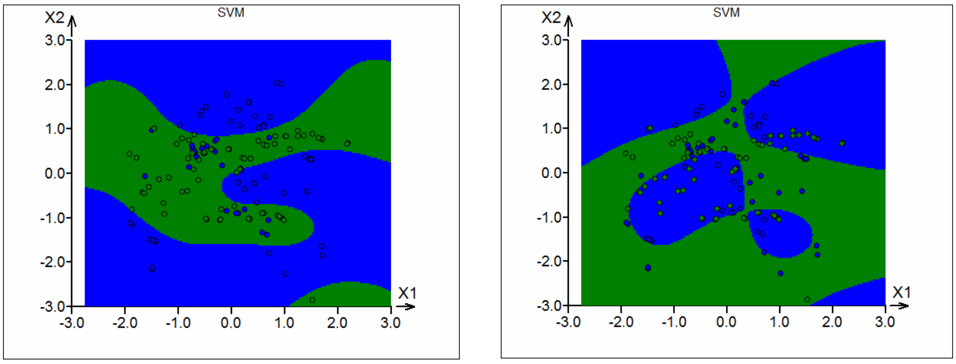

Figure 6: SVM classification model on projection in only the first two principal component in the transformed parameter space. Misclassification rate about 15 – 20%

Table 1: Typical classification results for a series from two technological states from the same gauge

svm_type	c_svc		Misclass Table	
kernel_type	RBF		Correct	151
Gamma	0.5		Misclass	2
cases	153		Misclass Rate	0.013
variables	5			
nr_class	2			
total_sv	49			
label	1	2		
nr_sv	26	23		

4. Conclusion

Behaviour an applicability of the feedforward neural networks in industrial technological applications on time series models was investigated. It was shown that although the solution and parameters of the ANN model are often overdetermined and unstable, reducing dimension with a set of bootstrapped optimal parameter vectors using classical principal component analysis may dramatically stabilise the found solutions and make them usable for prediction of important states, discrimination between predefined risks, forecast possible semi-random changes in technology and thus increase stability and reliability in technology and quality management. SVM classifier proved to be a very suitable tool thanks to its ability to discriminate between linearly non-separable and non-convex subsets.

References

1. K. Kupka, M. Meloun: The end-point estimation in instrumental titrations by nonlinear regression, Analytica Chimica Acta, 429 (1), p.171-183, Feb 2001.
2. E. Alpaydin, M. I. Jordan: Local linear perceptrons for classification. IEEE Transactions on Neural Networks, 7(3) May 1996.
3. Stefan Ruping: SVM Kernels for Time Series Analysis, CS Dept, University of Dortmund, http://www-ai.cs.uni-dortmund.de/EVENTS/FGML2001/ FGML2001-Paper-Rueping.pdf
4. Sayan Mukherjee, Edgar Osuna, Federico Girosi: Nonlinear Prediction of Chaotic Time Series Using Support Vector Machines, IEEE Workshop on Neural Networks for Signal Processing VII (1997).
5. W. A. C. Schmidt, J. P. Davis: Pattern Recognition Properties of Various Feature Spaces for Higher Order Neural Networks, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 15 (8) 1993, 795 – 801.
6. G. E. Hinton, R. R. Salakhutdinov: Reducing the Dimensionality of Data with Neural Networks, Science, Vol 313, July 2006, 504-507.
7. Patterson D W, Chan K H, Tan C M.: Time Series Forecasting with neural nets: a comparative study. Proc. International Conference on Neural Network Applictions to Signal Processing, NNASP 1993 Singapore pp 269-274.
8. K. Muller, A. Smola, G. Ratsch, B. Scholkopf, J. Kohlmorgen, and V. Vapnik: Predicting time series with support vector machines. In Proceedings of the International Conference on Artificial Neural Networks, Springer Lecture Notes in Computer Science. Springer, 1997.
9. C. L. Giles, S. Lawrence, A. C. Tsoi: Noisy Time Series Prediction using Recurrent Neural Networks and Grammatical Inference, Machine Learning, 44, 161–183, 2001.
10. Zhang Q, Benveniste A: (1992) Wavelet Networks. IEEE Trans. Neural Netw. 3: 889-898.
11. V. Vapnik: Statistical Learning Theory. Wiley, Chichester, GB, 1998.
12. Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin: A Practical Guide to Support Vector Classification, Department of Computer Science, National Taiwan University, Taipei 106, Taiwan, http://www.csie.ntu.edu.tw/~cjlin
13. K. Kupka: QCExpert 3.2, User Manual – Advanced Statistical Methods, TriloByte Statistical Software, 2009, http://www.trilobyte.cz