Feed-Forward and Recurrent Neural Networks in Signal Prediction

Ales Prochazka and Ales Pavelka
Institute of Chemical Technology in Prague, Department of Computing and Control Engineering
Technicka Street 5, 166 28 Prague 6, Czech Republic
Phone: +420 220 444 198 * Fax: +420 220 445 053 * Email: A.Prochazka@ieee.org * Web: http://dsp.vscht.cz

***Abstract*—The paper is devoted to time series prediction using linear, perceptron and Elman neural networks of the proposed pattern structure. Signal wavelet de-noising in the initial stage is discussed as well. The main part of the paper is devoted to the comparison of different models of time series prediction. The proposed algorithm is applied to the real signal representing gas consumption.**

***Index Terms*—AR modelling, neural networks, Elman networks, signal prediction, distributed computing**

I. INTRODUCTION

Signal analysis and prediction [1], [2] are very important tools used in a wide range of engineering applications [3]. The paper presents selected linear and non-linear methods of signal prediction [4], [5], [6], [7] including both linear models, feed-forward neural networks and recurrent structures used for energy consumption forecasting [8], [9], [10], [11].

The general mathematical description is followed by the processing of the given signal of gas consumption in the Czech Republic presented in Fig. 1. Signal modelling is verified using both original and de-noised values obtained by wavelet signal decomposition and thresholding. The corresponding spectrum estimation points to periodic components of the observed signal and provides basic information about the model structure selection.

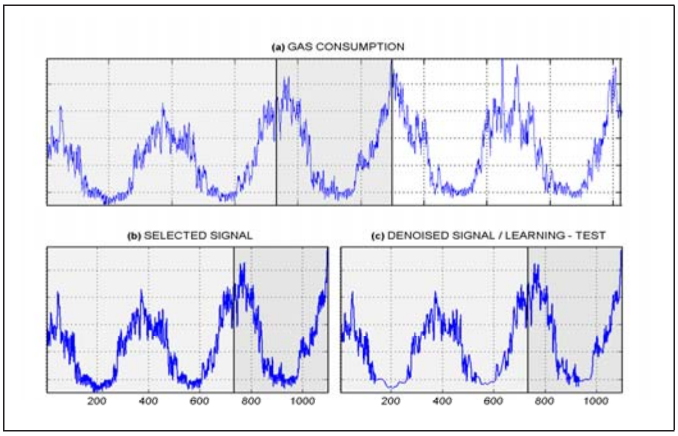

Fig. 1. Gas consumption in the Czech Republic **(a)** during the period 2001-05 observed with the sampling period of one day, **(b)** the period of two years used for prediction and one year for verification, and **(c)** observed values after wavelet rejection of signal noise parts

Thanks to Research grant No. MSM 6046137306

Proposed algorithms have been verified in the Matlab mathematical environment using distributed computing owing to the large amount of computations needed to optimize neural network coefficients.

II. SIGNAL WAVELET DE-NOISING

Wavelet functions use in signal prediction allow signal de-noising, multi-resolution forecasting and signal reconstruction. The initial (*mother*) wavelet $w(t)$ modified by dilation $a = 2^m$ and translation $b = k\ 2^m$ forms the set of functions

$$W_{m,k}(t) = \frac{1}{\sqrt{a}}\, w(\frac{t-b}{a}) = \frac{1}{\sqrt{2^m}}\, w(2^{-m}\ t - k) \quad (1)$$

for integer values m, k used for signal decomposition [12].

Signal de-nosing can then be done by appropriate thresholding of wavelet coefficients according to Fig. 2. In the case of soft-thresholding it is possible to evaluate new coefficients $\overline{c}(k)$ using original coefficients $c(k)$ for a chosen threshold value δ by relation

$$\overline{c}(k) = \begin{cases} \text{sign } c(k)\ (\mid c(k) \mid -\delta) & \text{if } \mid c(k) \mid > \delta \\ 0 & \text{if } \mid c(k) \mid \leq \delta \end{cases} \quad (2)$$

Results of de-nosing of a real signal segment using local thresholding in Fig. 2 present original and de-noised signals and corresponding wavelet coefficients.

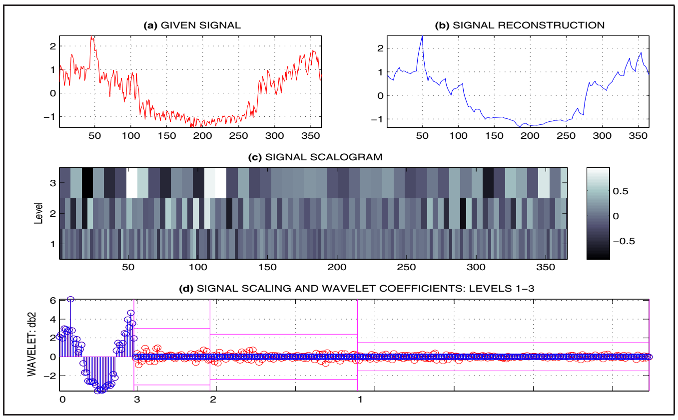

Fig. 2. Wavelet de-noising of the signal segment presenting **(a)** original signal, **(b)** de-noised signal, **(c)** signal scalogram, and **(d)** wavelet coefficients resulting from the decomposition into the third level and their local thresholding

III. PREDICTION MODELS

Basic model structures for signal prediction are presented in Fig. 3. All these models assume block-oriented processing evaluating model coefficients in the learning part for the given pattern matrix and using this model in the verification part.

In all these cases both the original signal $\{x(n)\}$ and its wavelet de-noised version is used to study the effect of signal preprocessing to the quality of signal prediction.

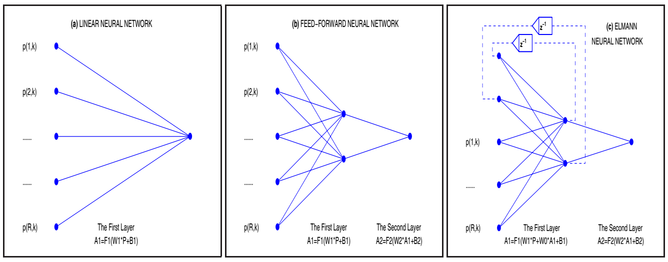

Fig. 3. Fundamental models for signal prediction including **(a)** a linear model, **(b)** feed-forward, and **(c)** Elmann neural network

A. Autoregressive model

The autoregressive (AR) model used for prediction of a given signal $\{x(n)\}$ can be defined by the linear neural network according to Fig. 3(a) for vector of coefficients

$$\mathbf{W} = \begin{pmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,R} \end{pmatrix} \tag{3}$$

the pattern matrix

$$\mathbf{P}=\begin{pmatrix} p_{1,1} \cdots p_{1,k} \cdots p_{1,Q} \\ p_{2,1} \cdots p_{2,k} \cdots p_{2,Q} \\ \cdots \quad \cdots \\ p_{R,1} \cdots p_{R,k} \cdots p_{R,Q} \end{pmatrix}, \; \mathbf{p}_k=\begin{pmatrix} p_{1,k} \\ p_{2,k} \\ \cdots \\ p_{R,k} \end{pmatrix}=\begin{pmatrix} x_{R+k-1} \\ x_{R+k-2} \\ \cdots \\ x_k \end{pmatrix} \tag{4}$$

and target values

$$\mathbf{T} = \begin{pmatrix} t_{1,1} & \cdots & t_{1,k} & \cdots & t_{1,Q} \end{pmatrix}, \qquad t_{1,k}=x_{R+k} \tag{5}$$

Defining vector of evaluated values as

$$\mathbf{A} = \begin{pmatrix} a_{1,1} & \cdots & a_{1,k} & \cdots & a_{1,Q} \end{pmatrix} \tag{6}$$

with its $k-th$ element defined by relation

$$a_{1,k} = \mathbf{W} * \mathbf{p}_k = \sum_{j=1}^{N} w_{1,j}\, p_{j,k} \tag{7}$$

it is possible to use the criterium function

$$S(W) = \sum_{j=1}^{Q} (a_{1,k} - t_{1,k})^2 = \sum_{j=1}^{Q} (w_{1,j}\, p_{j,k} - t_{1,k})^2 \tag{8}$$

to find values $\{w_{1,j}\}$ by the least square method. Owing to the linearity this problem results in the system of R linear algebraic equations to find model coefficients $\{w_{1,j}\}$.

The first step in signal modelling includes the selection of structure of the pattern matrix and the estimation the model order. Having the initial matrix $\mathbf{P}$ we can find its subset containing its selected rows only [1] having the most significant contribution to signal prediction.

The selection procedure is based upon the singular value decomposition [13] of the pattern matrix into the product of three matrices and observation of the distribution of singular values to choose the reduced model order. The QRp factorization is then applied to select the most significant model coefficients i.e. those which have the most important influence to the accuracy of the model.

B. Perceptron model

The classical perceptron model of the tructure $R-S1-S2$ illustrated in Fig. 3(b) includes two matrices of coefficients

$$\mathbf{W1} = \begin{pmatrix} w1_{1,1} & w1_{1,2} & \cdots & w1_{1,R} \\ & & \cdots & \\ w1_{S1,1} & w1_{S1,2} & \cdots & w1_{S1,R} \end{pmatrix} \tag{9}$$

and

$$\mathbf{W2} = \begin{pmatrix} w2_{1,1} & w2_{1,2} & \cdots & w2_{1,S1} \\ & & \cdots & \\ w2_{S2,1} & w2_{S2,2} & \cdots & w2_{S2,R} \end{pmatrix} \tag{10}$$

Using the pattern matrix (4) it is possible to evaluate network output (6) by relations

$$\mathbf{A1} = F1(\mathbf{W1} * \mathbf{P} + \mathbf{B1}) \tag{11}$$

$$\mathbf{A} = F2(\mathbf{W2} * \mathbf{A1} + \mathbf{B2}) \tag{12}$$

where $\mathbf{B1}, \mathbf{B2}$ stand for biases and $F1, F2$ represent transfer functions. In the case of one signal values prediction (S2=1) the same criterium as that described by Eq. (8) can be used.

C. Elman model

Elman network standing for the recurrent neural with the structure presented in Fig. 3(c) is based on the optimization of two matrices of coefficients summarized in Eq. (9) and (10) using Eq. (11) and the pattern matrix

$$\mathbf{P} = \begin{pmatrix} \cdots & p_{1,k} & \cdots \\ & \cdots & \\ \cdots & p_{S1,k} & \cdots \\ \cdots & p_{S1+1,k} & \cdots \\ \cdots & p_{S1+2,k} & \cdots \\ & \cdots & \\ \cdots & p_{R,k} & \cdots \end{pmatrix} \equiv \begin{pmatrix} a1_{1,k-1} \\ \cdots \\ a1_{S1,k-1} \\ x_{R+k-1} \\ x_{R+k-2} \\ \cdots \\ x_k \end{pmatrix} \tag{13}$$

for zero initial conditions and target values

$$\mathbf{T} = \begin{pmatrix} \cdots & t_{1,k} & \cdots \end{pmatrix} \equiv \begin{pmatrix} \cdots & x_{R+k} & \cdots \end{pmatrix} \tag{14}$$

Results of Elman learning applied to the de-noised signal of gas consumption are presented in Fig. 4.

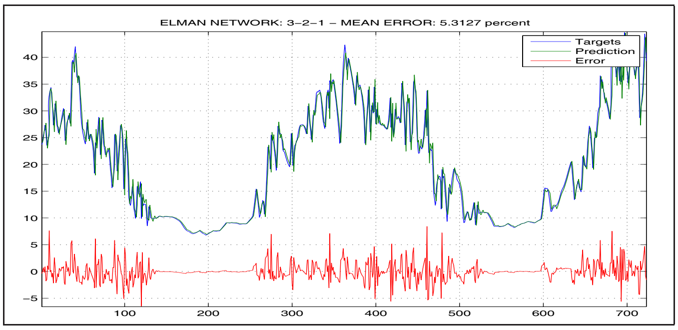

Fig. 4. Results of Elman learning for the data of gas consumption using wavelet de-noising presenting given and predicted signals and their difference

IV. RESULTS

Models described above have been applied for processing od gas consumption in the Czech Republic using both original and de-noised signals. Measured values of data consumption and temperature evolution were observed with the sampling period of one day.

Owing to the large amount of computations all numerical tests were performed using distributed computing with 8 computers (workers) in a cluster and typical computational time of 2-3 hours for 16 numerical experiments (tasks) specifying 2-3 years for the learning process and one year for model verification. Fig. 5 presents a typical job report for all computers contributing to the job processing. Results achieved present the mean percentage error for linear, feed-forward and Elman neural network for each task and their average. It is possible to find the positive effect of the network recurrence.

Figs 6 and 7 compare results achieved in the learning and verification parts. The structure of pattern values includes the use of several values from signal history in each pattern vector selected according to results of spectral analysis.

Experimental values obtained are summarized in Table I and II. Errors achieved in both these parts are of the same order and it is possible to see how Elmann networks are able to decrease this error. Wavelet transform de-noising (WTdenoising) can further decrease the prediction error.

V. CONCLUSION

The paper presents an analysis of basic algorithms used for linear and non-linear signal prediction with numerical experiments of gas consumption prediction. Results include comparison of data prediction applied to original and de-noised values using wavelet coefficient thresholding as well.

All experiments were done using distributed computing owing to the large amount of computations. This approach allows an efficient data prediction. The pattern vectors structure has been selected according to signal spectrum analysis. It is assumed that this selection process will be improved by application of SVD decomposition and QR factorization.

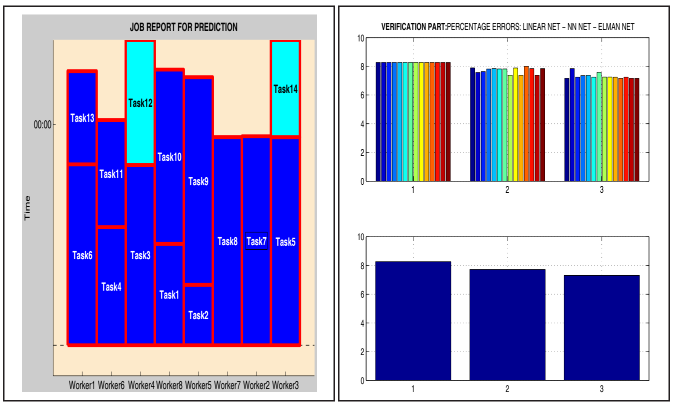

Fig. 5. Job report of distributed computing and percentage mean errors achieved in the learning signal part for each task

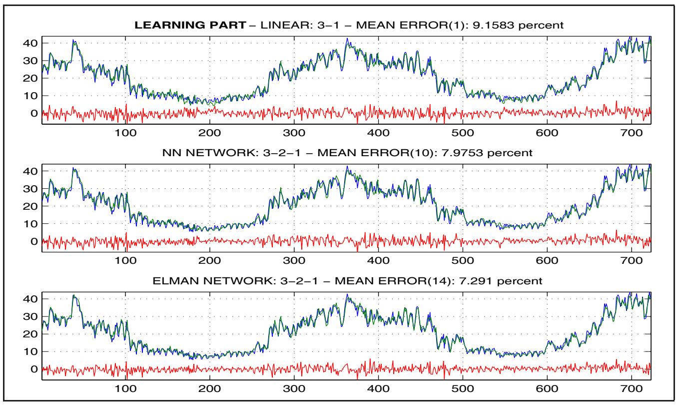

Fig. 6. Results the learning process for basic model structures of signal prediction including a linear model, feed-forward network, and Elman with corresponding errors

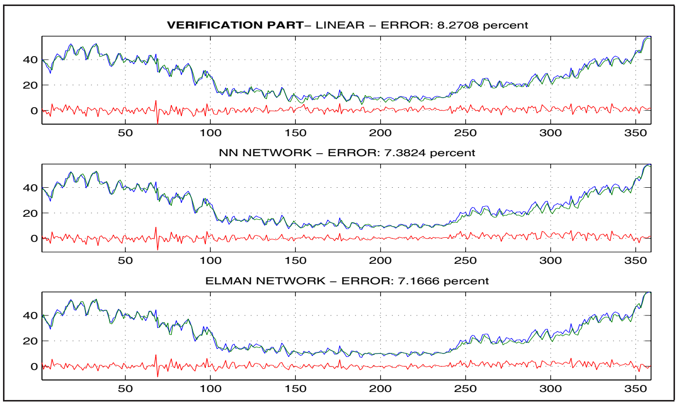

Fig. 7. Data verification using structures proposed in the learning part system optimization with corresponding errors

Further mathematical research will be devoted to analysis of block oriented signal prediction including the study of optimal block size selection and its comparison with real time neural network prediction allowing adaptive signal modelling with time-varying coefficients.

TABLE I

THE MEAN PERCENTAGE ERROR OF ONE STEP AHEAD PREDICTION BY LINEAR, PERCEPTRON AND RECURRENT NEURAL NETWORKS FOR THE PATTERN MATRIX HAVING 3 VALUES OF CONSUMPTION AND TWO OPTIONAL VALUES OF TEMPERATURE USING DAILY VALUES FROM THE TWO YEAR PERIOD FOR ONE YEAR AHEAD PREDICTION

		Mean Error [%]		
		Linear 3-1	*NN* 3-2-1	*Elman* 3-2-1
Patterns: Consumption only				
1 (99 min)	Learning Part	9.68	9.23	8.65
	Verification Part	8.42	8.48	7.84
2 (114 min)	Learning Part	5.55	5.15	4.23
WTdenoising	Verification Part	6.00	5.37	4.93
Patterns: Consumption - temperature (2 additional nodes)				
3 (113 min)	Learning Part	9.16	7.98	7.29
	Verification Part	8.27	7.38	7.17
4 (114 min)	Learning Part	6.31	4.89	4.09
WTdenoising	Verification Part	7.04	5.55	5.47

TABLE II

THE MEAN PERCENTAGE ERROR OF ONE STEP AHEAD PREDICTION BY LINEAR, PERCEPTRON AND RECURRENT NEURAL NETWORKS FOR THE PATTERN MATRIX HAVING 3 VALUES OF CONSUMPTION AND TWO OPTIONAL VALUES OF TEMPERATURE USING DAILY VALUES FROM THE THREE YEAR PERIOD FOR ONE YEAR AHEAD PREDICTION

Test		*Mean Error [%]*		
		Linear 3-1	*NN* 3-2-1	*Elman* 3-2-1
Patterns: Consumption only				
5 (216 min)	Learning Part	9.21	8.91	8.21
	Verification Part	8.16	8.14	7.76
Patterns: Consumption - temperature (2 additional nodes)				
6 (217 min)	Learning Part	8.75	7.72	6.97
	Verification Part	7.24	6.82	6.10

In order to study the relevance of an individual model to a given system it is necessary to find suitable signal pre-processing methods as well. In this connection further research will be devoted to the influence of de-noising [14] of input data using appropriate methods including the application of wavelet transforms for signal decomposition, subsequent thresholding and reconstruction.

It is assumed that further research will be also devoted to the appropriate selection and neural networks structure [15], [16] for signal prediction.

REFERENCES

[1] P. P. Kanjilal, *Adaptive Prediction and Predictive Control*, Peter Peregrinus Ltd., IEE, U.K., 1995.

[2] A. S. Weigend and N. A. Gershenfeld, Eds., *Time Series Prediction*, Addison-Wesley Publishing, 1994.

[3] J. W. Taylor, L. M. de Menezes, and P. E. McSharry, "A comparison of univariate methods for forecasting electricity demand up to a day ahead," *International Journal of Forecasting*, vol. 22, no. 1, pp. 1–16, 2006.

[4] B.Q. Huang, Tarik Rashid, and M.-T. Kechadi, "Multi-context recurrent neural network for time series applications," *International Journal of Computational Intelligence*, vol. 3, no. 1, 2006.

[5] Tarik Rashid, B.Q. Huang, M-T. Kechadi, and Brian Gleeson, "Auto regressive recurrent neural network approach for electricity load forecasting," *International Journal of Computational Intelligence*, vol. 3, no. 1, 2006.

[6] T. Koskela, M. Lehtokangas, J. Saarinen, and K. Kaski, "Time Series Prediction with Multilayer Perceptron, FIR and Elman Neural Networks," in *Proceedings of the World Congress on Neural Networks, INNS Press*, 1996, pp. 491–496.

[7] D. Benaouda and F. Murtagh, "Neuro-wavelet approach to time-series signals prediction: An example of electricity load and pool-price data," *International Journal of Emerging Electric Power Systems*, vol. 8, no. 2, 2007.

[8] Tarik Rashid, M.-T. Kechadii, and B.Q. Huang, "Short-term energy load forecasting using recurrent neural network," in *The 8th IASTED International Conference on artificial intelligence and soft computing*, Marbella, Spain, September 2004, vol. 1, pp. 451–150.

[9] Tarik Rashid and M-T. Kechadi, "Study of artificial neural networks for daily peak electricity load forecasting," The 2nd International Conference on Information Technology (ICIT 2005) Amman, Jordan, 2005.

[10] Tarik Rashid, M-T.Kechadi, and Brian Gleeson, "Neural network based methods for electric load forecasting," The 3rd International Conference on Cybernetics and Information Technologies, Systems and Applications (CITSA 2006) in Orlando, USA, 2006.

[11] D. P. Mandic and J. A. Chambers, *Recurrent Neural Networks for Prediction*, John Wiley & Sons, Chichester, U.K., 2001.

[12] D. E. Newland, *An Introduction to Random Vibrations, Spectral and Wavelet Analysis*, Longman Scientific & Technical, Essex, U.K., third edition, 1994.

[13] A. Procházka, M. Mudrová, and M. Štorek, "Wavelet Use for Noise Rejection and Signal Modelling," in *Signal Analysis and Prediction*, A. Procházka, J. Uhlíř, P. J. W. Rayner, and N. G. Kingsbury, Eds., chapter 16. Birkhauser, Boston, U.S.A., 1998.

[14] D. L. Donoho, "De-Noising by Soft-Thresholding," *IEEE Trans. on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.

[15] M. C. Medeiros, T. Terasvirta, and G. Rech, "Building neural network models for time series: a statistical approach," *Journal of Forecasting*, vol. 25, no. 1, pp. 49–75, 2006.

[16] Hanh Nguyen and Christine Chan, "Multiple neural networks for a long term time series forecast ," *Neural Computing & Applications*, vol. 13, no. 4, pp. 90–98, 2004.