ARTIFICIAL NEURAL NETWORKS IN PREDICTION AND PREDICTIVE CONTROL

David Samek
Dept. of Production Engineering
Tomas Bata University in Zlin
Mostni 5139, 76001 Zlin
E-mail: samek@ft.utb.cz

Petr Dostal
Department of Process Control
Tomas Bata University in Zlin
Nad Stranemi 4511, 76005 Zlin
E-mail: dostalp@fai.utb.cz

KEYWORDS
Artificial neural networks, prediction, model predictive control.

ABSTRACT

In this contribution the three various artificial neural networks are tested on CATS prediction benchmark. The results are compared and evaluated. Furthermore, these artificial neural networks are tested in model predictive control on the t-variant system. The aim of this paper is to present and compare artificial neural networks as interesting way how to model and predict nonlinear systems even with t-variant parameters. The key features of this paper are emphasis of the computational costs of the selected predictors and usage of adaptive linear network which offers short learning times and remarkable prediction error.

INTRODUCTION

The increasing demand on the quality, reliability, and economical profits leads to using of new modeling and control methods in the process industry. In past few decades the predictive control techniques have become very popular. One of the most used approaches is the Model Predictive Control (MPC) method (Camacho and Bordons 1995).

The appropriate predictive model is a key question in nonlinear model predictive control. The predictive models can be divided into two main groups (Verdunen and Jong 2003): white box models and black box models. The white box modeling is established on a prior knowledge of mathematic description of basic physical rules of controlled process. White box models are excellent for process modeling and product development. The model constants have a physical meaning and are not dependent on process design. The main disadvantage of white box models is the time of development and higher complexity. Conversely, black box models such as artificial neural network (ANN) and fuzzy logic models are data-driven. They provide general method for process dynamics description from input-output data. First and foremost, the learning ability makes artificial neural networks versatile, user friendly and powerful tool for many practical applications (Hussain 1999).

Many predictive control techniques based on MPC, which use artificial neural network as a predictor, are established on multilayer feed-forward neural networks (Hagan et al. 2002; Kanjilal 1995). In spite of the fact that the multilayer feed-forward neural networks (MFFNNs) have many advantages, such as simple design and scalability, they have also many drawbacks, such as long training times and choice of an appropriate learning stop time (the over-learning versus the early stopping). Nevertheless, there are quite a number of ANN types suitable for the modeling and prediction (Liu 2001; Meszaros et al. 1999; Chu et al. 2003). Moreover, features of these ANNs exceed abilities of the MFFNN in many cases. One of these ANNs is ADALINE (ADAptive LInear NEuron). What is more, ADALINE has one special feature – adaptivity. Owing to its simple structure it offers interesting way how to design adaptive neural predictor with reasonable computational demands. This paper is organized as follows: In the beginning multilayer feed-forward neural networks and adaptive linear networks are briefly introduced. Then the methodology of the simulations is explained, after that the results are presented and the paper is concluded by final remarks.

MULTILAYER FEED-FORWARD NEURAL NETWORKS

Multilayer feed-forward neural networks were derived by generalization from Rosenblatt's perceptron, thus they are often called multilayer perceptrons (MLP). This type of artificial neural networks uses supervised training. One of the most known methods of supervised training is backpropagation algorithm; hence these ANNs are sometimes also called backpropagation networks.

In the MFFNN the signals flow between the neurons only in the forward direction i.e. towards the output. Neurons in MFFNN are organized in layers and neurons of the certain layer can have inputs from any neurons of the earlier layer. The ability to predict of ANN is determined by capability of modeling of certain process. By applying the Kolmogorov theorem it was proved that for general function approximation is sufficient two-layer MFFNN (one hidden layer) if non-polynomial transfer functions are used and the hidden layer has enough neurons (Leshno et al. 1993).

Proceedings 22nd European Conference on Modelling and Simulation ©ECMS Loucas S. Louca, Yiorgos Chrysanthou, Zuzana Oplatková, Khalid Al-Begain (Editors)
ISBN: 978-0-9553018-5-8 / ISBN: 978-0-9553018-6-5 (CD)

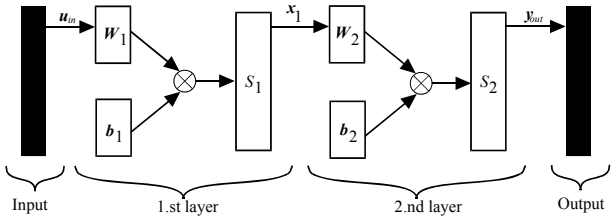

Figure 1: Simplified Scheme of Two-layer MFFNN

The two-layer MFFNN, which contains one output layer and one hidden layer, is depicted in the figure 1 (this structure is implemented in this paper). This MFFNN can be described by two equations:

$$\boldsymbol{x_1} = S_1(\boldsymbol{b_1} + \boldsymbol{W_1} \cdot \boldsymbol{u_{in}}) \quad (1)$$

$$\boldsymbol{y_{out}} = S_2(\boldsymbol{b_2} + \boldsymbol{W_2} \cdot \boldsymbol{x_1}) \quad (2)$$

Where $\boldsymbol{y_{out}}$ is the network output vector, S_i is transfer function of i-th layer, $\boldsymbol{b_i}$ is bias vector of i-th layer, $\boldsymbol{W_i}$ is weighting matrix of i-th layer, $\boldsymbol{x_1}$ is output vector of the first layer and $\boldsymbol{u_{in}}$ is the network input vector.

ADAPTIVE LINEAR NETWORKS

ADALINE was described by B. Widrow and M.E. Hoff as an adaptive threshold logic element in 1960 (Widrow and Hoff 1960). Though, the original version of ADALINE had only simple two-state threshold transfer function with the range of function {-1;+1}, nowadays ADALINE is also used with linear transfer function (Arbib 2002; Demuth and Beale 2002). Although this structure has limited skills, it is possible to connect more of ADALINEs together to obtain a MADALINE (Multiple ADALINE).

Despite the fact that ADALINEs are able to solve only linearly separable problems, it has been shown in practice that they can approximate nonlinear functions with sufficient accuracy while using enough number of neurons (Freeman and Skapura 1991).

Because of their main advantage, that is very fast learning, they have many practical applications, e.g. noise reduction, signal processing and signal prediction in control and communication systems. Simplified scheme of ADALINE is shown in figure 2.

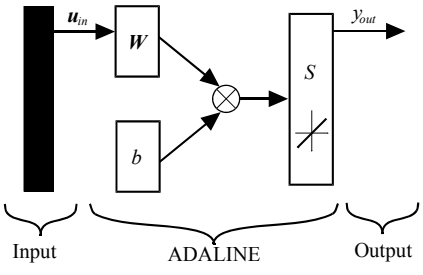

Figure 2: Simplified Scheme of ADALINE with Linear Transfer Function

The learning procedure is based on an iterative search process, where performance feedback is used to guide the search process. In other words, a designer "trains" the system by "showing" it examples of inputs and the respective outputs. In this way, system competence is directly and quantitatively related to the amount of experience the system was given. The most popular learning method is simple LMS (Least Mean Square) algorithm (Widrow and Hoff 1960), often called the Widrow Hoff Delta Rule (Rumelhart et al. 1986), which is adopted in this paper. This method is based on the minimization of Mean Square Error (MSE).

METHODOLOGY

The three different artificial neural networks were tested in prediction problem. There were chosen two different variants of two-layer feed-forward neural network. The first structure followed from Kolmogorov theorem (Leshno et al. 1993) and had hyperbolic tangent transfer function in both layers (this structure will be in the following text denoted as *mffnntt*), while the second structure uses very common layout with hyperbolic tangent transfer function in hidden layer and linear transfer function in output layer (this structure will be in the following text denoted as *mffnntp*). The third predictor was based on ADALINE (this structure will be in the following text denoted as *adaline*).

The artificial neural networks were tested in two tasks. The first task was the prediction of time series defined by CATS benchmark (Vandewalle et al. 2007; Lendasse et al. 2004). Then the artificial neural networks were used as the predictors of Model Predictive Controller.

CATS Benchmark

The CATS (Competition on Artificial Time Series) benchmark was designed in order to compare various prediction methods. The task of the predictor is to predict 100 missing values of the 5000 artificial time series data. The missing values are divided in 5 blocks – 981 to 1,000; 1,981 to 2,000; 2,981 to 3,000; 3,981 to 4,000; 4,981 to 5,000.

CATS benchmark uses two criterions E_1 and E_2:

$$E_1 = \frac{\sum_{t=981}^{1000}(e_t - \hat{e}_t)^2}{100} + \frac{\sum_{t=1981}^{2000}(e_t - \hat{e}_t)^2}{100} + \frac{\sum_{t=2981}^{3000}(e_t - \hat{e}_t)^2}{100} + \frac{\sum_{t=3981}^{4000}(e_t - \hat{e}_t)^2}{100} + \frac{\sum_{t=4981}^{5000}(e_t - \hat{e}_t)^2}{100} \quad (3)$$

$$E_2 = \frac{\sum_{t=981}^{1000}(e-\hat{e})^2}{80} + \frac{\sum_{t=1981}^{2000}(e-\hat{e})^2}{80} + \frac{\sum_{t=2981}^{3000}(e-\hat{e})^2}{80} + \frac{\sum_{t=3981}^{4000}(e-\hat{e})^2}{80} \quad (4)$$

Where e is the real value of the signal, $\hat{e}$ is the predicted value and t is the time step.

Model Predictive Control

The selected artificial networks were tested in the modeling and prediction of nonlinear system of two interconnected ball tanks. Involving usual simplifications, mathematical model of the system can be defined:

$$\pi h_1(d_1 - h_1)\frac{dh_1}{dt} + q_1 = q_{1v} \quad (5)$$

$$\pi h_2(d_2 - h_2)\frac{dh_2}{dt} - q_1 + q_2 = q_{2v} \quad (6)$$

Where d_j are tanks diameters, h_j are levels of liquid in the tanks, q_j are output flow rates and q_{jv} are input flow rates (j denotes number of tank). Input flow rates depend on the liquid levels:

$$q_1 = k_1\sqrt{|h_1 - h_2|} \quad (7)$$

$$\text{if } h_1 - h_2 < 0 \text{ then } q_1 = -q_1 \quad (8)$$

$$q_2 = k_2\sqrt{h_2} \quad (9)$$

Where k_1 and k_2 are constants describing the pipes. Initial conditions of equations (5) and (6), that were obtained from steady state, are $h_1(0)$=1,5m, $h_2(0)$=1,3m, $q_{1v}(0)$=0.38m^3/min, $q_{2v}(0)$=0.19m^3/min. System is regarded as SISO – controllers task is to control liquid level in the first tank (h_1=y) by influencing input flow rate q_{1v} (u). Second input flow rate remains constant.

In order to test the adaptivity of the predictors, parameters d_1 and d_2 were rapidly changed during simulation at time t=100min. This change is physically unfeasible and can be done only in the simulation. These parameters were chosen with the intention of better demonstration of the adaptivity feature.

The controller uses classical MPC objective function (Camacho and Bordons 1995):

$$J = \sum_{i=N_1}^{N_2}\left[w(k+i) - \hat{y}(k+i)\right]^2 + \rho \cdot \sum_{i=1}^{N_u}\left[u_t(k+i-1) - u_t(k+i-2)\right]^2 \quad (10)$$

where N_1, N_2 and N_u define horizons over which the tracking error and the control increments are evaluated (usually $N_2 \geq Nu$). The u_t variable is the tentative control signal, w is the desired response and $\hat{y}$ is the network model response. The parameter ρ determines the contribution that the sum of the squares of the control increments has on the performance index.

There is usually assumed that after a certain interval $N_u < N_2$ there is no variation in the proposed control signals, that is:

$$\Delta u(k+i) = 0 \quad \text{for} \quad i \in \langle N_u, N_2 - 1\rangle \quad (11)$$

This is equivalent to giving infinite weights to the changes in the control from a certain instant. This approach is adopted in this paper.

In addition to visual comparisons of control courses, quadratic criterions were also used to evaluate individual control courses. Two quadratic criterions were used to describe behavior of individual controllers.

$$S_y = \sum_{j=0}^{2000}\left(w_j - y_j\right)^2 \quad (12)$$

$$S_u = \sum_{j=0}^{2000}\left(u_j - u_{j-1}\right)^2 \quad (13)$$

The symbol w_j stands for the desired value of the controlled variable (the level of liquid) in the step j, y_j is the controlled variable in the step j, u_i is the control signal in the step j. The limits j=0 and j=2000 corresponds to the simulation time 0 min and 200 min (because of the sampling rate 0,1 min).

The S_y criterion is based on the control error and thus represents the tracking performance of the controller. The criterion S_u is based on changes of control signal and represents the controller demands on the actuators.

SIMULATIONS AND RESULTS

All simulations were performed using MATLAB/Simulink.

CATS Benchmark

By many experiments it was discovered that ideal (sufficient) structure for MFFNN is 5 neurons in input layer (zero layer), 10 neurons in hidden layer and 1 output neuron. Furthermore, it was found out that number of training epochs has key influence on the prediction quality. Thus, it was decided to explore this influence. There were created, trained and tested 100 networks for each of ten different maximum numbers of training epochs (MTE) from the interval 25-250. In the tables 1-3 are presented averages of criterions E_1, E_2, final global error FGE, real number of training epochs RTE, prediction time PT and train time TT.

As well as in case MFFNN, ADALINE has important training parameter – learning rate (LR). The influence of this parameter was observed in the interval 10^{-2}-10^{-11}. And, again, it was done 100 simulations for each value of learning rate. The average values of criterions are presented in the table 3. The structure of *adaline* contained 5 neurons in the input (zero) layer and 1 neuron in the output layer.

Table 1: CATS Benchmark of the *mffnntt*

MTE	E_1 (10^4)	E_2 (10^4)	FGE (10^{-4})	RTE	PE (s)	TT (s)
25	3,02	2,93	52,65	25,00	0,41	7,04
50	2,44	2,44	66,92	50,00	0,41	14,07
75	2,17	2,12	10,55	75,00	0,41	21,11
100	1,76	1,61	6,89	100,00	0,41	28,26
125	1,60	1,51	5,98	124,92	0,41	35,13
150	1,49	1,38	5,70	149,69	0,41	42,06
175	1,48	1,38	5,54	171,19	0,41	48,05
200	1,46	1,34	5,43	191,82	0,41	53,73
225	1,46	1,34	5,34	208,67	0,41	58,68
250	1,52	1,41	5,32	219,60	0,41	62,17

Table 2: CATS Benchmark of the *mffnntp*

MTE	E_1 (10^4)	E_2 (10^4)	FGE (10^{-4})	RTE	PE (s)	TT (s)
25	55,61	55,27	34,56	25,00	0,40	7,28
50	9,98	10,62	13,66	50,00	0,40	14,10
75	5,80	6,05	8,67	75,00	0,40	21,11
100	2,37	2,42	7,08	100,00	0,40	28,02
125	3,51	3,48	6,36	124,75	0,40	34,88
150	4,06	3,93	5,76	149,13	0,40	41,72
175	1,69	1,61	5,40	173,58	0,40	48,51
200	2,21	2,17	5,28	195,21	0,40	54,52
225	2,22	2,16	5,14	213,43	0,40	59,11
250	2,02	2,02	5,12	225,35	0,40	62,24

As can be seen from tables 1-3, the best results from the point of view of criterions E_1 and E_2 were obtained by *mffnntt*. The *mffnntp* has slightly worse results, but not very significantly. The highest values of the criterions were obtained by *adaline*. However, the computational times (training time TT and prediction time PT) were shortest in case of *adaline*.

As can be seen from table 1, the best CATS predictor from the tested group is *mffnntt* trained approximately 200 epochs. From the table 3 it can be concluded that it is useless to set lower learning rate (LR) than $1 \cdot 10^{-4}$.

Table 3: CATS Benchmark of the *adaline*

LR	E_1	E_2	FGE	RTE	PE (s)	TT (s)
10^{-2}	$7{,}59 \cdot 10^{42}$	$8{,}97 \cdot 10^{42}$	0,17	1	0,55	0,13
10^{-3}	$4{,}99 \cdot 10^{13}$	$5{,}75 \cdot 10^{13}$	0,17	1	0,62	0,16
10^{-4}	$2{,}50 \cdot 10^{4}$	$2{,}66 \cdot 10^{4}$	0,17	1	0,51	0,09
10^{-5}	$2{,}50 \cdot 10^{4}$	$2{,}46 \cdot 10^{4}$	0,17	1	0,55	0,14
10^{-6}	$2{,}51 \cdot 10^{4}$	$2{,}45 \cdot 10^{4}$	0,17	1	0,69	0,21
10^{-7}	$2{,}51 \cdot 10^{4}$	$2{,}45 \cdot 10^{4}$	0,17	1	0,38	0,14
10^{-8}	$2{,}51 \cdot 10^{4}$	$2{,}45 \cdot 10^{4}$	0,17	1	0,39	0,16
10^{-9}	$2{,}51 \cdot 10^{4}$	$2{,}45 \cdot 10^{4}$	0,17	1	0,50	0,14
10^{-10}	$2{,}51 \cdot 10^{4}$	$2{,}45 \cdot 10^{4}$	0,17	1	0,46	0,11
10^{-11}	$2{,}51 \cdot 10^{4}$	$2{,}45 \cdot 10^{4}$	0,17	1	0,51	0,13

Model Predictive Control

Controller parameters were set to: ρ=0,8; N_1=1; N_2=20 and N_u=5. All predictors (ANNs) used 5 recent values of control signal and 5 recent values of output signal for prediction. Thus, the ANNs had ten inputs and one output.

The structure of *mffnntt* and *mffnntp* consisted 10 neurons in input layer (zero layer), 25 neurons in hidden layer and 1 output neuron. The structure of *adaline* contained 5 neurons in the input (zero) layer and 1 neuron in the output layer.

Due to satisfactory control performance during beginning of experiment all predictors were trained off-line from off-line prepared data. However, the ADALINE predictor was adapted during its function at each time step. Simulation results are shown in the figures 3-5 and table 4.Controller uses sampling rate 0,1 min and Levenberg-Marquart method was used as the optimization algorithm of the (10).

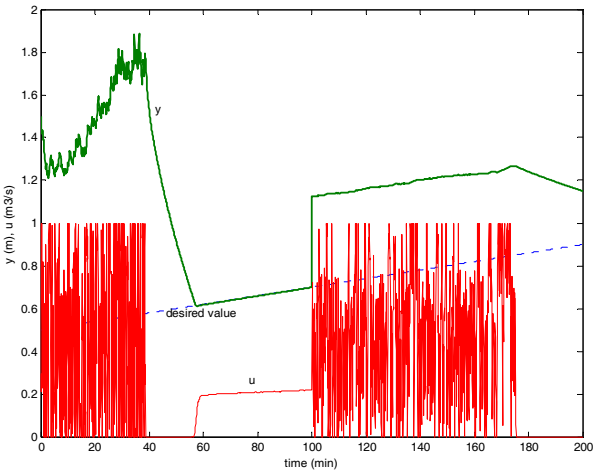

Figure 3: Control Using *mffnntt*

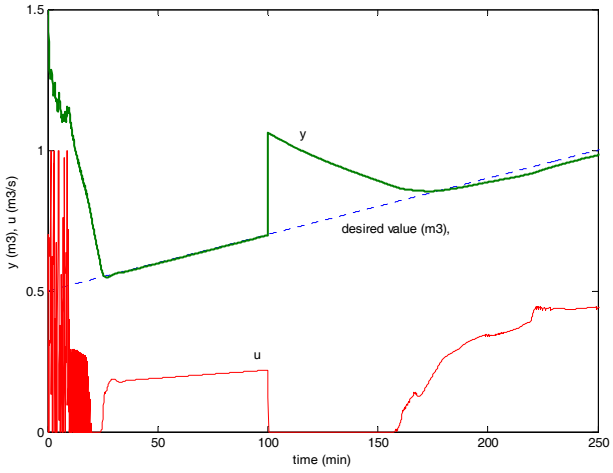

Figure 4: Control Using *mffnntp*

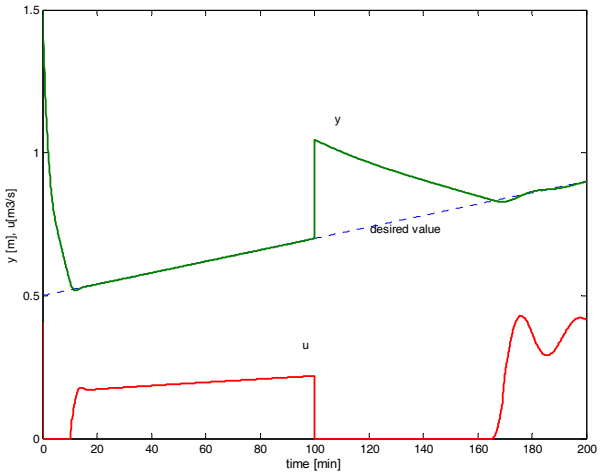

Figure 5: Control Using *adaline*

Table 4: Comparison of the Predictors in MPC

Predictor of the controller	Criterion S_y	Criterion S_u
mffnntt	577,39	140,51
mffnntp	92,84	15,22
adaline	40,23	0,22

As can be seen from figures 3-5 and the table 4, the best results were gained using *adaline* predictor. The predictors based on MFFNN lead to oscillations of control signal. This behavior could be reduced by changing controller parameters (especially increasing ρ parameter). Furthermore, from the courses after change of system parameters (time=100min) it can be concluded that MFFNN based predictors do not provide natural adaptive prediction. However, ADALINE from its structure offers adaptivity, as was depicted in the figure 5. What is more, ADALINE has very simple structure thus the computational demands of the controller are notably low.

CONCLUSION

From the simulations it can be stated that selection of suitable predictor based on artificial network must be done very carefully. For various predicted systems has to be used different predictors. The multilayer feed-forward neural networks provided best results in the CATS benchmark of predictors. Nevertheless, in the model predictive control with the same types of artificial neural networks MFFNN did not offered applicable results. On the other hand ADALINE in the control task showed impressive results, including control quality, fast training and prediction times and adaptivity.

ACKNOWLEDGEMENT

This work has been supported by the Ministry of Education, Youth and Sports of the Czech Republic under grant MSM 7088352102 and by the Grant Agency of the Czech Republic under grant 102/07/P137. This support is greatly acknowledged.

REFERENCES

Arbib, M. A. 2002.*The Handbook of Brain Theory and Neural Networks*. The MIT Press, Cambridge, MA.

Camacho, E.F. and C. Bordons. 1995. *Model Predictive Control in the Process Industry*. Springer – Verlag, Berlin.

Chu J.Z.; Tsai, P.F.; Tsai, W.Y.; Jang, S.S.; Wong, D.S.H.; Shieh, S.S.; Lin, P.H.; and S.J. Jiang. 2003. "An Experimental Study of Model Predictive Control Based on Artificial Neural Networks". In *Proceedings of 7th International Conference on Knowledge-Based Intelligent Information and Engineering Systems* (Oxford, UK, Sep 3-5), Springer, Oxford, UK, 1296-1302.

Demuth, H. and M. Beale. 2002. *Neural Network Toolbox for Use with MATLAB*. The Mathworks, Inc., Natick, MA.

Freeman, J.A. and D.M. Skapura. 1991. *Neural Networks Algorithms, Applications, and Programming Techniques*. Addison Wesley Publishing Company, Reading, MA.

Hagan, M.; Demuth H.; and O.D. Jesus. 2002. "An Introduction to the Use of Neural Networks in Control Systems." *International Journal of Robust and Nonlinear Control*, No.12 (Nov), 959-985.

Hussain, M.A. 1999. "Review of the application of neural networks in chemical process control – simulation and online implementation." *Artificial Intelligence in Engineering*, No.1 (Jan), 55-68.

Kanjilal, P.P. 1995. *Adaptive prediction and predictive control*. P. Perengrinus Ltd., London.

Lendasse, A.; Oja, E.; and O. Simula. 2004. "Time Series Prediction Competition: The CATS Benchmark". In *International Joint Conference on Neural Networks* (Budapest, Jul 25-29), IEEE, 1615-1620.

Leshno, M.; Lin, V.Y.; Pinkus, A.; and S. Schocken. 1993. "Multilayer feedforward networks with a non-polynomial activation can approximate any function." *Neural networks*, No. 6, 861-867.

Liu, G.P. 2001. *Nonlinear identification and control – A neural network approach*. Springer - Verlag London Ltd., London.

Meszaros, A.; Rusnak, A.; and M. Fikar. 1999. "Adaptive Neural PID Control – Case Study: Tubular Chemical Reactor," *Computers and Chemical Engineering Supplement*, No.1999, 847-850.

Rumelhart, D. E.; Hinton, G. E.; and R. J. Williams. 1986. "Learning internal representations by error propagation". In *Parallel distributed processing: explorations in the microstructure of cognition, vol. 1: foundations*, D. E. Rumelhart and J. L. McClelland (Eds.). MIT Press, Cambridge, MA, 318-362.

Vandewalle, J. Suykens, J. De Moor, B. and A. Lendasse. 2007. "State-of-the-art and Evolution in Public Data Sets and Competitions for System Identification, Time Series Prediction and Pattern Recognition". In *32nd International Conference on Acoustics, Speech, and Signal Processing* (Honolulu, HI, Apr 15-20). Hawaii Convention Center in Honolulu, HI, 1269-1272.

Verdurmen R. E. M. and P. de Jong. 2003. "Optimising product quality and process control for powdered dairy products". In *Dairy processing: Improving quality*. G. Smith (Ed.). Woodhead Publishing Limited, Cambridge, 333-365.

Widrow B. and M. E. Hoff, Jr. 1960. "Adaptive switching circuits". In *IRE WESCON Convention Rec.* (New York). IRE, New York, 96–104.

AUTHOR BIOGRAPHIES

DAVID SAMEK was born in Zlin, Czech Republic and went to the Brno University of Technology, where he studied automation and control and obtained his degree in 2000. Then he studied post gradual in Tomas Bata University in Zlin and obtained Ph.D. degree in 2005. Since 2001 he has been working at Tomas Bata University in Zlin as a senior lecturer. His field of interest is focused on artificial networks and their applications. His e-mail address is : `samek@ft.utb.cz` and his Web-page can be found at `http://www.ft.utb.cz/czech/uvi/`.

PETR DOSTAL studied at Institute of Chemical Technology in Pardubice, where he studied automation of chemical processes and obtained his masters degree in 1968 and Ph.D. degree in 1976. He has been working Nowadays he works as a professor and head of Department of Process Control in Tomas Bata University, Faculty of Applied Informatics. His research is focused on field of interest is focused on artificial networks and their applications. His e-mail address is : `dostalp@fai.utb.cz` and his Web-page can be found at `http://www.fai.utb.cz/`.