Multilayer Feed-Forward Neural Networks in Prediction and Predictive Control of Semi-Batch Reactor

Lubomir Macku and David Samek

Abstract—The contribution studies prediction of the given semi-batch reactor using multilayer feed-forward neural networks. The two prediction approaches are tested — signal prediction approach and system prediction methodology. The first approach is commonly applied in time series prediction, while the input-output models in the second methodology are used for example in the control tasks. Furthermore, the resulting predictor is used for the model predictive control of the reactor in order to test performance of the developed method.

Keywords— Artificial neural network, chemical reactor, prediction, predictive control.

I. INTRODUCTION

Prediction of nonlinear and complex systems can be performed by various methods. One of them is simplification and linearization that leads to linear models [1, 2]. Some authors use wavelet filtering in order to divide stochastic and deterministic parts that are modeled separately [4, 5]. Javadi et al. and Phaiboon present successful application of fuzzy logic to prediction [6, 7]. However, probably the most popular approach in these cases is based on artificial neural networks.

Artificial neural networks are commonly used in various fields, for example weather forecasting [8], time series prediction of financial data [9, 10], biology and medicine [11, 12], power engineering [13] and process control [14, 15]. There is lot of types of artificial neural networks, but not all of them are usable for prediction. The most common are multilayer feed-forward neural networks [8, 11, 16, 17]. Fairly wide group of artificial neural networks belongs to recurrent neural networks [18, 19]. Very popular due to their fast training are radial basis function neural networks [18, 19]. Al-Shayea and El-Refae use generalized regression neural network for prediction of Spanish banks data [16].

Manuscript received July 30, 2012: Revised version received August 16, 2012.

L. Macku is with the Department of Electronics and Measurements, Faculty of Applied Informatics, Tomas Bata University in Zlin, nam. T. G. Masaryka 5555, 760 01 Zlin Czech Republic (e-mail: macku@fai.utb.cz).

D. Samek is with the Department of Production Engineering, Faculty of Technology, Tomas Bata University in Zlin, nam. T. G. Masaryka 5555, 760 01 Zlin Czech Republic (phone: +420-576-035-157; fax: +420-576-035-176; e-mail: samek@fb.utb.cz).

also other less frequent methods such as fuzzy-neural networks [20], adaptive linear networks [21], unsupervised Kohonen neural networks [22, 23]. Interesting generalization of neural networks, which is achieved by using multi-argument and learnable functions, bring functional networks [24].

Generally, it can be distinguished between two main approaches in the prediction. The first approach deals with the predicted signal only. It means that the predicted value must be predicted from known older values of the same signal. This approach is very common in financial engineering and it is usually denoted as a signal prediction or time series prediction (in case of time series).

The second approach uses other signals (usually system inputs) for the prediction and it is commonly called system or process prediction. This approach is widely used in the process industry, especially in the control (predictive control).

In this contribution both approaches to prediction are studied and the multilayer feed-forward neural networks are tested on the prediction of temperature in the chemical semi-batch reactor. The motivation of this work comes from our long time research focused on predictive control of chemical reactors. Usually, the predictor must be adapted during the function, because of its inaccuracy or changing parameters of the system to be controlled. The goal of the presented work is to design and test the proper predictor based on multilayer feed-forward neural network that would not require adaptation during control of the given reactor.

The paper is structured along these lines: First, the chemical semi-batch reactor is introduced. After that, artificial multilayer feed-forward neural networks are explained briefly. The following part describes the predictors design, their testing and validation. Then, the resulting predictors are compared and discussed. This chapter is followed by the explanation of the model predictive control of the given semi-batch reactor. The final part of the paper contains few concluding remarks.

II. CHEMICAL SEMI-BATCH REACTOR

In this paper, a semi-batch reactor model is used to study the prediction abilities of multilayer feed-forward neural networks and their applicability in the model predictive control. The model input data comes from a real process - the chromium waste recycling process [15]. The task of the predictor is to...
predict the in-reactor temperature.

Let us consider a single input – single output (SISO) system of the chemical exothermic semi-batch reactor, which is described by the following mathematical model:

$$
\frac{dm(t)}{dt} = F_t
$$

(1)

$$
\frac{da(t)}{dt} = \frac{F_i}{m(t)} - A e^{\frac{-E}{R T(t)}} a(t)
$$

(2)

$$
\frac{dT(t)}{dt} = \frac{F_i \cdot c_l \cdot T_i}{m(t) \cdot c} + \frac{A e^{\frac{-E}{R T(t)}} \cdot \Delta H_r \cdot a(t)}{c}
- \frac{K \cdot S \cdot T(t)}{m(t) \cdot c} + \frac{K \cdot S \cdot T_C(t)}{m(t) \cdot c}
$$

(3)

$$
\frac{dT_C(t)}{dt} = \frac{F_C \cdot T_{Cl}}{m_C} + \frac{K \cdot S \cdot T(t)}{m_C \cdot c_C}
- \frac{K \cdot S \cdot T_C(t)}{m_C \cdot c_C} - \frac{F_C \cdot T_C(t)}{m_C}
$$

(4)

where \(m\) is the total weight of the reaction components in the reactor, \(a\) is the mass concentration of the reaction component in the reactor, \(c\) = 4500 J kg K\(^{-1}\) is the specific heat capacity of the reactor content, \(T\) is the temperature of the reactor content. \(F_t, T_j = 293.15\) K and \(c_l = 4400\) J kg K\(^{-1}\) is the reaction component input mass flow rate, the temperature and the specific heat capacity. \(F_C = 1\) kg s\(^{-1}\), \(T_{Cl} = 288.15\) K, \(T_C, c_C = 4118\) J kg K\(^{-1}\) and \(m_C = 220\) kg is the cooling water mass flow rate, the input temperature, the output temperature, the specific heat capacity and the weight of the cooling water in the cooling system of the reactor, respectively.

Other constants: \(A = 219.588\) s\(^{-1}\), \(E = 29967.5087\) J mol\(^{-1}\), \(R = 8.314\) J mol\(^{-1}\) K\(^{-1}\), \(\Delta H_r = 1392350\) J kg\(^{-1}\), \(K = 200\) kg s\(^{-1}\) K\(^{-1}\) and \(S = 7.36\) m\(^2\) is the effective heat-transfer area [15].

III. MULTILAYER FEED-FORWARD NEURAL NETWORKS

Multilayer feed-forward neural networks (MFFNNs) are very often called backpropagation networks because of the typical training algorithm. Some authors prefer name multilayer perceptrons (MLP) [25, 26], because MFFNNs have been developed by generalization from Rosenblatt’s perceptron with binary transfer function. This structure is depicted in the Fig. 1. The \(u_m\) stands for input value, the \(N\) is number of inputs and \(y\) is the output of the Rosenblatt’s perceptron.

In MFFNN the signals flow only in the one direction (from the input to the output). All neurons are structured into layers and, typically, all neurons in the specific layer use same transfer function. In contrast to Rosenblatt’s perceptron multilayer feed-forward neural networks use various transfer functions, usually continuous (e.g. linear, hyperbolic tangent, sigmoidal functions, etc.). Example of the two-layer feed-forward neural network is shown in the Fig. 2. The \(u_m\) is the input data vector, \(W\) is weighting matrix, \(b\) stands for bias vector, \(S\) is transfer function and \(x\) is output of internal layer.

By applying of Kolmogorov theorem it has been proved [27] that two-layer MFFNN (with one hidden layer) can approximate any function with certain accuracy while non-polynomial transfer function in the hidden layer is used. This methodology is adopted in the paper.

IV. PREDICTION

As was already mention hereinafter, there were studied two approaches of the prediction: signal prediction and system prediction. For all simulations in this paper MATLAB with Neural Network Toolbox, Optimization Toolbox and Simulink were used.

In the first case the artificial neural network used five past values of the predicted signal as the input vector and predicted only one step ahead. In other words, when it was needed the MFFNN repeatedly used its own predictions as inputs. Thus, in the input (zero) layer of the artificial network were five neurons and the output layer consisted one neuron.
As is depicted in the Fig. 3, the network had five neurons in the hidden layer. This number was obtained by many experiments as sufficient for this case. The hidden layer used hyperbolic tangent sigmoid as the transfer function, while the output layer employed linear transfer function.

Three simulation results from our previous work [15] were chosen as the training and testing data. Only one simulation data were used for training (furthermore they will be symbolized as data1), whilst the rest two simulation results served as testing group (data2 and data3). The training was performed using built-in Matlab/Neural Network Toolbox Levenberg-Marquardt backpropagation algorithm. This artificial neural network will be in the following text denoted as net1.

After the network training, the obtained predictor was tested to all three signals (Fig. 5, 7, 9). For better comparability the prediction error for all testing signals was computed (Fig 6, 8, 10).

In order to compare influence of the input signal knowledge to the prediction quality, the second predictor was also tested. This predictor used not only the predicted signal old values but also old values of the system input. In this case it was the chromium sludge dosing speed. Thus the multilayer feedforward neural network involved ten neurons in the input layer (see Fig. 4). As well as in the previous approach, the hidden layer used hyperbolic tangent sigmoid as the transfer function, while the output layer employed linear transfer function. This artificial neural network will be in the following text denoted as net2.

With the intention of obtaining comparable results the same methodology as in case of net1 was used. It means that same training data, the same training algorithm and same testing data were applied. Results are depicted in Fig. 11-16.
Fig. 8 Prediction error of net1 for the testing data (data2)

Fig. 9 Net1 test on the testing data (data3)

Fig. 10 Prediction error of net1 for the testing data (data3)

Fig. 11 Net2 test on the training data (data1)

Fig. 12 Prediction error of net2 for the training data (data1)

Fig. 13 Net2 test on the testing data (data2)
V. COMPARISON OF PREDICTORS

For better comparison two prediction quality criterions were defined. The first criterion $C_1$ describes total sum of absolute values of prediction errors relative to number predictions, whilst the second criterion function $C_2$ characterizes total sum of squares of prediction errors relative to number predictions. The $C_1$ criterion gives same importance to all errors. On the other hand $C_2$ emphasizes higher errors and lower prediction errors are suppressed.

$$C_1 = \frac{\sum_{i=1}^{N} \left| t(i) - p(i) \right|}{N}$$  \hspace{2cm} (5)

$$C_2 = \frac{\sum_{i=1}^{N} (t(i) - p(i))^2}{N}$$  \hspace{2cm} (6)

$N$ stands for number of predictions (length of the predicted signal), $t$ is target (original) signal, $p$ denotes predicted signal and $i$ is number of the prediction.

Table I. Prediction quality criterions

<table>
<thead>
<tr>
<th></th>
<th>net1</th>
<th>net2</th>
<th>net1</th>
<th>net2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$C_1$</td>
<td>$C_2$</td>
<td>$C_1$</td>
<td>$C_2$</td>
</tr>
<tr>
<td>data1</td>
<td>0.0024</td>
<td>4.529e-05</td>
<td>9.828e-05</td>
<td>2.218e-06</td>
</tr>
<tr>
<td>data2</td>
<td>0.0133</td>
<td>0.0013</td>
<td>0.0115</td>
<td>0.0107</td>
</tr>
<tr>
<td>data3</td>
<td>0.0565</td>
<td>0.0806</td>
<td>0.0155</td>
<td>0.0080</td>
</tr>
</tbody>
</table>

As can be seen from the results, the prediction is in both cases very good. However, if the best method should be judged, the second approach ($net2$) must be chosen, because it provided much better results with the one and only exception (criterion $C_2$ for $data2$). Therefore, this predictor will be tested in the model predictive controller of the given reactor in the following text.

VI. MODEL PREDICTIVE CONTROL

The selected predictor $net2$ was applied into model predictive controller (MPC) introduced in our previous paper [15]. The controller uses explicit predictor and optimization box based on the Levenberg-Marquart method. The criterion function $J$ to be optimized is defined along these lines:

$$J = \lambda \sum_{j=1}^{N} \left[ y_r(k+j) - \hat{y}(k+j) \right]^2 +$$

$$+ \rho \sum_{j=1}^{N} \left[ u_r(k+j-1) - u_r(k+j-2) \right]^2 +$$

$$+ \gamma(k) \sum_{j=1}^{N} u_r(k+j)$$

$$\gamma(k) = \gamma(k-1) - \gamma_c$$  \hspace{2cm} (8)

where the $\lambda$, $\rho$ and $\gamma$ parameters determine the contribution that the particular sum has on the performance index. The $\gamma$ parameter is decreasing during the control. The speed of the change is defined by the (8) using parameter $\gamma_c$. In other
words, the third sum in the beginning of the control has the maximum value, and after initial phase it equals to zero. The \( N_1, N_2 \) and \( N_u \) define horizons over which the tracking error and the control increments are evaluated. The \( u_t \) variable is the tentative control signal; \( y_r \) is the desired response and \( \hat{y} \) is the predicted value of the temperature. Index \( k \) symbolizes step of the control (sample time of the control was 60s).

The settings of the controller are adopted from the [15] in order to obtain comparable result (see Table II). These settings will be in the following text denoted as \( \text{controller1} \). However, with the intention of revealing the \( \text{net2 predictor abilities} \) the other controller settings were tested. The all controllers used same horizons \( N_2 = N_u = 8 \) and parameter \( N_1 = 1 \).

As can be seen from Fig. 17, the \( \text{net2 predictor based} \) MPC controller successfully controls temperature in the given chemical reactor. The dosing time was 3060 s and maximum in-reactor temperature reached 370.5 K. The decreasing temperature of the reaction mixture after the time 3060 s is caused by the fact that the reactor works in semi-batch cycle. In other words, after consumption of the all dosing batch (chromium sludge), the control is finished and the operator must wait until the reactor content cools down. The Fig. 18 shows course of the control signal (the dosing speed of the chromium sludge). The Fig. 19 depicts the development of the total mass of the in-reactor mixture. The speed of the reaction is illustrated in the Fig. 20, where can be seen steep changes of the chromium sludge concentration. The temperature of the cooling water in the cooling system was also observed in our simulations (Fig. 21).

Table II. MPC controller settings

<table>
<thead>
<tr>
<th>Controller</th>
<th>( \lambda )</th>
<th>( \rho )</th>
<th>( \gamma )</th>
<th>( \gamma_c )</th>
</tr>
</thead>
<tbody>
<tr>
<td>\text{controller1}</td>
<td>1000</td>
<td>100000</td>
<td>10000</td>
<td>200</td>
</tr>
<tr>
<td>\text{controller2}</td>
<td>1000</td>
<td>100000</td>
<td>10000</td>
<td>100</td>
</tr>
<tr>
<td>\text{controller3}</td>
<td>1000</td>
<td>1</td>
<td>600000</td>
<td>10000</td>
</tr>
<tr>
<td>\text{controller4}</td>
<td>1000</td>
<td>100000</td>
<td>600000</td>
<td>10000</td>
</tr>
<tr>
<td>\text{controller5}</td>
<td>1000</td>
<td>100000</td>
<td>600000</td>
<td>18000</td>
</tr>
</tbody>
</table>

Fig. 17 The in-reactor temperature obtained using \text{controller1}

Fig. 18 The chromium-sludge dosing obtained using \text{controller1}

Fig. 19 The mass of the reaction mixture obtained using \text{controller1}

Fig. 20 The in-reactor chromium sludge concentration obtained using \text{controller1}
Fig. 21 The temperature in the cooling system obtained using controller1

Fig. 22 The in-reactor temperature obtained using controller2

Fig. 23 The chromium-sludge dosing obtained using controller2

Fig. 24 The in-reactor temperature obtained using controller3

Fig. 25 The chromium-sludge dosing obtained using controller3

Fig. 26 The in-reactor temperature obtained using controller4
This behavior needs to be removed, but it leads to longer dosing times (Fig. 24 - 29). The controller3 had the dosing time 3180 s and oscillating control signal $F_i$ (Fig. 24 - 25). By increasing the $\rho$ parameter of the controller the control signal becomes smoother, however the temperature is still fluctuating around the reference value and the dosing times are still longer (Fig. 26 and 29).

VII. CONCLUSION

This paper continues and advances our previous work published in [28] and [15]. The artificial neural network based predictor developed in this paper offers comparable results as three-layer feed forward neural network used earlier. The proposed predictor is simpler and faster with equivalent accuracy. The simulations of the various settings of the model predictive controller showed that even good and exact predictor may suffer when the controller is not properly set.

REFERENCES


