

STEEL PLANT MODELING AND ANALYSIS USING A NEURAL NETWORK

Accepted for ANNIE '99, Intelligent Engineering Systems Through Artificial Neural Networks, Volume ??, pp. ??, November 1999, St. - Louis, USA.

JÜRGEN VAN GORP

*Department ELEC, Vrije Universiteit Brussel
Pleinlaan 2, B-1050 Brussel, Belgium
Email: Jurgen.Van.Gorp@vub.ac.be*

ABSTRACT

Even if parts of a chemical process are well known and can be described precisely, describing a complete chemical plant is a difficult task. In this paper a steel plant is modeled that can be seen as a nonlinear MIMO system. The control of the plant is currently based on human experience. A model of the plant is asked in order to optimize the steel quality and plant throughput. The problems that have to be overcome are a limited number of measurement samples, low accuracy and a high uncertainty. Different Neural Network models are described in order to make an optimal use of the given measurements. Two novel techniques are used in this paper. First an optimal plant model is assumed based on the results of a large number of NN models and early stopping. Second a proposal for optimization of the plant is given by introducing a novel method that uses Variance Backpropagation of the controlled output to obtain the input variances that can influence the product quality.

I. INTRODUCTION

The steel plant that is modeled in this paper is a chemical converter that burns down the carbon level in a mix of melted iron and chemical additives. This is done by blowing pure oxygen in the melt. The goal of the modeling effort is to predict the amount of pure oxygen that is needed in the process to reach a specified Carbon concentration and end temperature of the steel. In this paper the modeling is restricted to the prediction of the end temperature for an externally specified amount of oxygen. The underlying idea is that it is possible to control the plant, once it can be accurately modeled (Certainty Equivalence Principle [4]). We assume that the plant is invertible because to date this is done manually.

This paper is structured as follows. The steel plant is described and the measurement data is analysed in section II. The next section of the paper describes how a Neural Network is chosen and trained as a black box model for the steel plant. It is shown that the NN has a better prediction performance than the currently used manual predictions, and that an upper limit exists for the performance. Section IV gives an analysis of the measurement data, based on the obtained NN. Suggestions are given where to improve the measurements in order to improve the NN performance.

II. PLANT DESCRIPTION

The steel is processed in two phases. In a first stage raw iron and waste steel are melted with added chemicals. The melted iron is then poured in a converter and during the second stage pure oxygen is blown into the melt to obtain a specified Carbon content and specified goal temperature. This two stage process will further be called a *batch*. The total number of input and output parameters for the system equals 59. Not all of these parameters are useful for the NN modeling and a selection was made with a maximum of 25 input parameters, based on expert knowledge. Table 1 gives an overview of these parameters, with their tolerances and minimum/maximum boundaries. An (I) indicates that the value is known prior to the batch and can be used as an input to the NN. An (O) indicates that the value is measured at the end of the batch and can only be used for validation. The sensitivity values s_i^l and s_i^n are discussed in detail in section IV. In the sequel references to the parameters in table 1 will be written in *italic*.

The goal of the modeling is the prediction of *Phase2temp* within the boundaries [-10,+15] Kelvin from a given goal temperature. Denote h as the hitrate that the end temperature lays within this temperature window, or:

$$h = 100 \cdot P(\text{Phase2temp} \in \text{GoalTemp} + [-10, 15]) \quad (1)$$

To date, the hitrate for the manually controlled plant equals $h = 60.7\%$.

Parameter	[Min/ Max]	Tolerance	Unit	(I)input/ (O)utput	s_i^l	s_i^n
Converter age	[1/3000]			(I)		
Cooling time	[5/25]	1	min	(I)	0.32	0.30
Measure time	[1/4]	1	min	(I)	-1.75	-2.33
Batch type	[30/40]			(I)	0.22	0.2
Raw iron	[95/120]	0.36	ton	(I)	-1.21	-1.2
Added steel 1	[15/45]	?	ton	(I)	-2.89	-3
Added steel 2	[0/19]	?	ton	(I)	-3.39	-3.68
Added steel 3	[0/11]	?	ton	(I)	-4.73	-4.73
Start value C	[4/4.7]	2%	%	(I)	-8.55	-8.57
Goal value C	[0.03/0.06]	10%	%	(O)		
Goal temp.	[1650/1690]	0	K	(I)		
Oxygen	[7000/8500]	0.5%	m ³	(I)	33.010 ⁻³	33.310 ⁻³
Feed type 1	[800/1800]	10	mm	(I)	15.610 ⁻³	16.010 ⁻³
Feed type 2	[500/2000]	10	mm	(I)	3.9510 ⁻³	410 ⁻³
Feed type 3	[-150/150]	10	mm	(I)	-13.210 ⁻³	16.610 ⁻³
Mn	[0.4/1.4]	1%	%	(I)	8.23	9
Si	[0.3/1.7]	2%	%	(I)	38.5	35.7
N ₂	[0/40]	0.5%	ton	(I)	0.158	0.15
Ar	[0/60]	0.5%	ton	(I)	-43.110 ⁻³	-66.710 ⁻³
CaO ₂	[5/12]	0.05	ton	(I)	-5.6210 ⁻³	-5.710 ⁻³
Additive 1	[0/500]	10	kg	(I)		
Additive 2	[12/24]	5%	kg	(O)		
Additive 3	[3.5/9]	?	kg	(O)		
Additive 4	[0.1/0.45]	3%	%	(O)		
Phase 1 temp.	[1503/1683]	4	K	(I)	0.221	0.222
Phase 2 temp.	[1913/1983]	4	K	(O)		

Table 1 Main converter parameters used. s_i^l and s_i^n are defined further in the paper.

Analysis of the measurement data

The data was already filtered for measurement errors. After this manual filtering of the data set, 2091 measurement samples remained that were considered to be useable for learning. Histograms show that not all parameter have Gaussian shaped noise distributions. From the data it is impossible to determine whether this is due to uncertainties, random noise or *Batch Type* specifications. For that reason all noise contributions to the measurement data are considered to be Gaussian shaped with zero mean. For some parameters the tolerance is estimated, as shown in Table 1. In section IV a more detailed analysis is made in order to know if such an estimation is justified.

The samples in the data set are not necessarily subsequent. Large gaps exist in the *Converter age*, during which special batches were performed that are not included in the measurements. Analysis showed that the hitrate during subsequent batches is significantly higher. The reason is that the batch responsables based the amount of needed *Oxygen* on the results of previous batches. It is therefore justified to consider NARX models [4] besides NFIR models. Limited data sets are created by sliding a window over the current data set and rejecting data points if not all of the data points within the window are subsequent. These limited data sets will further be called *windowed data sets*. Fig. 1 shows how windowing reduces the number of useable data samples which puts severe constraints on the number of parameters that can be optimized. This is shown more in detail in the following section.

Windowing is also used to calculate variances on the measurement samples for use with Bayesian Learning [1] [5]. Calculating the variance of the data samples in a window must be done with caution. This is certainly the case with smaller window sizes, in which the smaller number of measurement samples can lead to unpredictable variances. For that reason the dynamic range of the variances is limited to 10dB.

Window Size	1	2	3	4	5	6	7
# Samples	2091	945	579	380	243	152	100

Fig. 1 Sample set size versus window size

Special care is needed when using the parameters *Raw Iron*, *Added Steel 1*, *2* and *3*. The amount of pure iron in these four parameters are guessed values and are fully based upon the skills of the batch responsables. E.g. the percentage of pure iron in a pile of raw ore is such an estimated value. In Table 2 the 95% boundaries of these guesses are shown, based on the mean of all guesses of the seven batch responsables. I.e. the mean of all guesses of each responsible is taken and the variance of the seven means is calculated. From the table can be concluded that these means have remarkable small 95% boundaries and that care must be taken when using these boundaries. For that reason 95% boundaries are taken, based on all measurements. The used boundaries therefore also include the variance on the used *Added Steel*, and not only measurement errors.

PARAMETER	MEAN VALUE	95% BOUNDARY (guessed)	95% BOUNDARY (measurements)
Added Steel 1	28.8 ton	+/- 1.22 ton	+/- 9.10 ton
Added Steel 2	7.18 ton	+/- 0.99 ton	+/- 7.17 ton
Added Steel 3	4.72 ton	+/- 0.57 ton	+/- 3.19 ton
Raw Iron	112.5 ton	+/- 0.36 ton	+/- 6.52 ton

Table 2 95% boundaries of guessed values

III. NEURAL NETWORK MAPPING

A. Used NN structure

The used Neural Network (NN) topology was a one hidden layer Multilayer Perceptron (MLP) with a tangent hyperbolic function in the hidden layer and a linear function for the output layer neuron. The number of hidden neurons was typically 5, while the number of neurons in the input and output layers differ with the number of chosen input and output parameters. The topology of the NN with 2091 measurement samples is shown in Fig. 2. The inputs are denoted as $x_k^{[i]}$ with $i = 1 \dots n$ and n the number of input parameters and $k = 1 \dots N$ with N the number of training samples. The outputs are denoted as $y_k^{[j]}$ with $j = 1 \dots m$ and m the number of output parameters. In this paper we consider only one output parameter (*Phase 2 temp*) such that $m = 1$. The weights for the first layer are defined by an weight matrix W_1 with size $5 \times (n + 1)$ (5 neurons and n inputs + bias term). For the output layer this becomes an $m \times 6$ matrix W_2 . The vector of NN parameters is denoted as $\theta = [W_1^T \ W_2^T]^T$. Remark that the number of parameters depends on the number of inputs and outputs of the system.

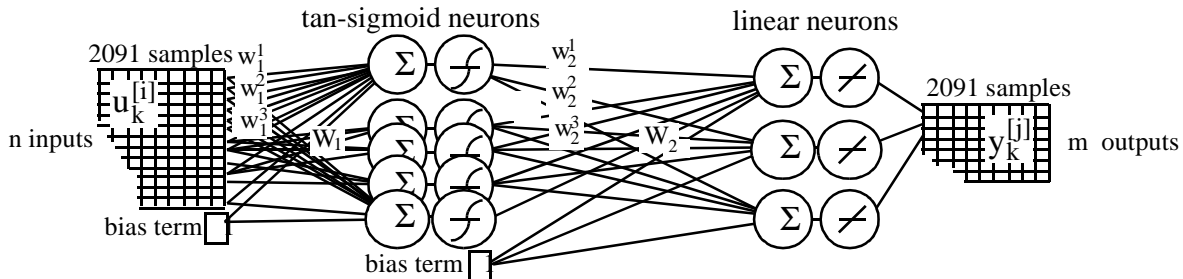


Fig. 2 Used NN topology

B. Choice of inputs and outputs

For the modeling, three different mappings were considered. The first was an NFIR structure

$$\hat{y}_k = f_{NN}(x_k, x_{k-1}, \dots, x_{k-p}, \theta) \quad (2)$$

with $x_k = [x_k^{[1]}, x_k^{[2]}, \dots, x_k^{[n]}]^T$, $y_k = [y_k^{[1]}, y_k^{[2]}, \dots, y_k^{[m]}]^T$ and in which $t = ws - 1$ and ws equals the window size. The first mapping used a Least Squares (LS) cost function, while the second mapping implemented the NFIR with a Bayesian cost function

$$J_B = \sum_{k=1}^N \frac{(y_k - f_{NN}(u_k, \theta))^2}{\sigma_y^2} \quad (3)$$

A third NN model mapped the dynamic behaviour with an NARX model (nonlinear ARX [2] [4])

$$\hat{y}_k = f_{NN}(x_k, x_{k-1}, \dots, x_{k-p}, y_{k-1}, y_{k-2}, \dots, y_{k-p}, \theta) \quad (4)$$

with an LS cost function. The NN function is defined as $\hat{y}_{NN} = W_2 \tanh(W_1(u_k))$ in which the u_k contains the

$x = [x_1, x_2, \dots, x_k, \dots]$ and $y = [y_1, y_2, \dots, y_k, \dots]$ parameters, enlarged with the bias term. The learning step was calculated using a Levenberg-Marquardt optimization step $\theta = -(J^T J + \lambda I)^{-1} J^T$, with $J = \partial E / \partial \theta$ the Jacobian matrix. The error vector E is defined as $E = y - f_{NN}(u, \theta)$ (for $m = 1$).

C. Identification results

Learning of the NN was done with the use of early stopping [3] in order to prevent overtraining. The data was split in three sets: one seventh of the data was used as a test set in order to test the performance of the obtained model. One fifth of the data was used as a validation set for early stopping. The remainder of the measurement data was used for learning. The sizes for test set and validation set were chosen arbitrarily, keeping in mind that only a limited amount of measurements was available such that choosing three sets of equal size was undesired. The low number of available measurements was a major problem. To fix the ideas consider the case where the NN was trained with 5 neurons, 21 inputs $u^{[i]}$ and one output $y^{[1]}$. The number of NN parameters (see also Fig. 2) for this case is 116. This becomes as high as 226 parameters if the full NARX model is used with all 43 inputs. The rule of thumb is that modeling should be done with at least 10 to 20 times more measurements than parameters. With 2,091 measurements this rule of thumb is met for both the simple NFIR and the NARX modeling. The *windowed* measurements however contain far less measurements (Fig. 1) and a trade-off must be made between the improved quality of the *windowed* samples and the complexity of the used model. Even if it is highly likely that the *windowed* measurements contain more information with respect to subsequent batches, the decrease in number of samples limited the use of the NARX model to a window size 2.

A total of 26 models was designed, with a different number of inputs and window sizes. Fig. 3 shows the mean performances of the models for the prediction of the end temperature of the steel, compared to the window sizes. The decrease in performance of prediction with larger window sizes is possibly due to the decrease in measurement data. From Fig. 3 it can be concluded that a prediction performance of 72% can be achieved with the NN modeling, compared to a hitrate of 60.7% that is currently reached by the batch leaders.

	Window Size			
	no window	2	3	4
NFIR (LS costfunction)	72%	72%	71%	69%
NFIR (Bayesian costfunction)	73%	73%	71%	69%
NARX	72%	74%		

Fig. 3 Performance of NN mappings versus window size

Another apparent conclusion that can be drawn from the NN mapping is that a performance above approximately 73% is not possible. All models were limited in performance by this hit rate. This indicates that the variance on the output is due to the variance of one of the inputs.

IV. PLANT ANALYSIS BASED ON THE NN IDENTIFICATION

A. Sensitivity Analysis

A sensitivity analysis is used to determine the importance of the different input parameters in the model. If the contribution of each parameter is known, conclusions can be drawn whether this parameter can be left out of the model. It might seem logical to include all possible inputs, no matter how big its contribution to the model is. Adding a single input to the model however increases the number of NN parameters with 5 (in the case of 5 hidden neurons). Since the number of data samples is limited, omitting unneeded inputs can lead to a better performing model.

Two numerical methods were considered for the sensitivity analysis. Both methods start from a mean input value, i.e. the system is expected to be excited by a very common input, that is based on mean values of each input. A first linear sensitivity analysis s_i^l calculates the first order partial derivative of the output with respect to each input, and is based on the local linearization of the NN model:

$$s_i^l = \frac{\partial}{\partial x^{[i]}} f_{NN}(\hat{x}, \theta), i = 1, 2, \dots, n \quad \text{with} \quad \hat{x} = \frac{1}{N} \sum_{k=1}^N x_k \quad (5)$$

A second nonlinear sensitivity analysis s_i^n also starts from the mean input values, but sweeps each input from its minimal value to its maximal value, while monitoring the NN mapping:

$$= \frac{(sup(f_{NN}(\hat{x}, \theta)) - inf(f_{NN}(\hat{x}, \theta)))|_{x^{[i]} \in [min(x^{[i]}, max(x^{[i]})]}]}{max(x^{[i]}) - min(x^{[i]})} \quad (6)$$

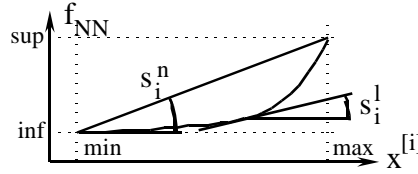


Fig. 4 Definitions of both sensitivity values

Care must be taken when interpreting the results from this sort of analysis and the results should not be extrapolated to any given measurement. Yet differences in both sensitivity values show to give a very good indication whether the contribution of a given input to the output temperature is linear or nonlinear. The s_i^l and s_i^n values are shown in Table 1. The table shows that all input parameters have a significant contribution and none should be left out.

B. Variance Analysis

From the overview of the results in Fig. 3 it is apparent that none of the models succeeds to perform better than a hitrate of approximately 73%. For that reason we state the following assumption.

Assumption 1 Due to the used early-stopping algorithm it is assumed that the NN mapping of the steel plant is near-optimal.

If assumption 1 is accepted, then the only reason for the variance on the output, is that the variance of one of the inputs is too large. Since the sensitivity of the output is known for each input, it is possible to calculate the output variance back to the inputs and compare them with the known or guessed input variances. This method has much correlation with the use of backpropagation for learning of the NN, where the error at the output is backpropagated through the NN to become the error on the NN parameters. A hitrate of 73% implies a standard deviation on the output of 11.05 degrees Kelvin, or a 95% interval ΔT_{steel} that equals $\Delta T_{steel} = 43.4$ ($Phase2temp \in [2194.3, 2237.7]$).

Parameter	$\Delta x^{[i]}$	$\Delta x^{[i]}$ guessed	s_i^n	$\Delta T_{steel}^{[i]}$	$\Delta T_{steel}^{[i]}$ guessed
Cooling time	1		0.3	0.3	
Measure time	1		2.33	2.33	
Batch type	0		0.2	0	
Raw iron	6.52	0.36	1.2	7.82	0.43
Added steel 1	9.10	1.22	3	27.3	3.66
Added steel 2	7.17	0.99	3.68	26.4	3.64
Added steel 3	3.19	6.52	4.73	15.1	2.70
Start value C	0.089		8.57	0.76	
Oxygen	38.51		33.310^{-3}	1.28	
Feed type 1	10		16.010^{-3}	0.16	
Feed type 2	10		4.010^{-3}	0.04	
Feed type 3	10		16.610^{-3}	0.17	
Mn	$7.48 \cdot 10^{-3}$		9	0.07	
Si	0.016		35.7	0.57	
N ₂	0.078		0.15	0.01	
Ar	0.044		66.710^{-3}	0.00	
CaO ₂	0.050		5.7110^{-3}	0.29	
Phase 1 temp.	4		.222	0.89	
Phase 2 temp.	4				

Table 3 Standard Deviation analysis

For backpropagation the interval for each parameter is calculated as

$$\Delta x_{backprop}^{[i]} = \Delta T_{steel} / s_i^n \quad (7)$$

and should be much larger than the actual interval $\Delta x^{[i]}$ of the parameter, if the variance of the parameters has no effect on the variance of the NN output. In order to clarify the results, we will not use backpropagation but instead use a forward calculation of the 95% interval $\Delta T_{steel}^{[i]}$ of each input as

$$\Delta T_{steel}^{[i]} = s_i^n \Delta x^{[i]}. \quad (8)$$

This means that, if a given input i has no influence on the variance of the NN, its $\Delta T_{steel}^{[i]}$ value should be significantly smaller than ΔT_{steel} . The analysis of Table 3 shows that the four previously mentioned parameters *Raw Iron*, *Added Steel 1*, *2* and *3* have 95% intervals that lay within an order of magnitude of ΔT_{steel} and that these inputs should be measured with a higher accuracy. In order to check the validity of the different $\Delta T_{steel}^{[i]}$ values, it is possible to calculate a theoretical 95% boundary for ΔT_{steel} , based on the measurements and knowing that

$$\Delta T_{steel, measurements} = \sqrt{\sum_i (\Delta T_{steel}^{[i]})^2} = 41.7 \quad (9)$$

which comes close to the true 95% interval ΔT_{steel} . From Table 3 we thus conclude that, in order to improve prediction on the steel end temperature, better measurements for *Added steel 1*, *2* and *3* are needed. Based on the guessed values, however, also the *Measure Time* and *Oxygen* values need better monitoring. These five values have the highest $\Delta T_{steel}^{[i]}$ intervals (>1).

V. CONCLUSIONS

In this paper a Neural Network was used for the prediction of the steel end temperature of a steel plant. Starting from a fixed number of prior measurements, the NN was able to predict the temperature with a hitrate of 72% within given boundaries. With this performance the currently used methods were outperformed by 12%. Starting from the assumption that this performance was the best that could be met with the given measurements, variance backpropagation was used to determine the inputs that limited the performance. Three input parameters could be appointed as possible causes for the output variance. In a next stage these three parameters should be measured with a higher accuracy in order to improve the nonlinear model of the steel plant and guarantee a better prediction.

VI. ACKNOWLEDGMENTS

This paper presents research results of the Belgian Programme on Interuniversity Poles of Attraction (IUAP 4/2), initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture; and the Flemish Government (GOA, IMMI2). The scientific responsibility rests with its authors.

VII. REFERENCES

- [1] C. M. Bishop, "Neural Networks for Pattern Recognition", Clarendon Press Oxford, 1994.
- [2] L. Ljung, "System Identification, Theory For The User", Prentice-Hall information and system sciences series, 1994.
- [3] J. Sjöberg, "Non-Linear System Identification with Neural Networks", Linköping Studies in Science and Technology, Dissertation No. 381, Linköping University, 1995.
- [4] J. A.K. Suykens et al., "Artificial Neural Networks for modeling and Control of Non-Linear Systems", Kluwer Academic Publishers, 1996.
- [5] J. Van Gorp, "The Errors-In-Variables Cost Function For Learning Neural Networks With Noisy Inputs", Intelligent Engineering through Artificial Neural Networks, Volume 8, Smart Engineering Systems, 1998, pp. 141 - 146.