In: Proceedings of the ICANN-93, Gielen S. and Kappen B. (eds.), Springer, Heidelberg, pp. 795-798

Estimating Material Properties for Process Optimization

Thomas Poppe and Thomas Martinetz
Siemens AG
Corporate Research and Development
Otto-Hahn-Ring 6
8000 München 83, Germany

A neural network approach to the problem of estimating physical properties of a material based on the material's chemical composition is presented. The network consists of sigmoidal hidden units and a linear output unit arranged in a feed-forward architecture. As a component of a process optimization system which is applied in production processes with a priori unknown and eventually drifting characteristics, robust and fast on-line adaptation of the network is required. Therefore, a permanently updated, stack-like organized training data set and a line-search procedure for adjusting the network weights is employed. A first application has been the estimation of the "relative yield stress" of different steel qualities, which is necessary for optimizing the rolling process at a hot line rolling mill. Compared to the current state-of-the-art method a reduction of the average estimation error of about 35% has been achieved.

1. Introduction

Process optimization requires knowledge about the relevant properties of the processed material. Depending on the material transformation process to be controlled, physical properties of the material like its heat capacity, its viscosity, its heat conductivity, or its hardness (just to mention a few) determine the optimal choice for the control parameter values. In most cases, however, the respective material property cannot be measured directly but must be estimated based on the thermodynamic state of the material, i.e., its chemical composition, its temperature, the given pressure, and eventually geometric quantities. The quality of the estimation result determines to a great extent the cost effectiveness and the product quality of the production process.

To be able to estimate material properties based on the thermodynamic state variables, the respective physical relationship has to be known. A common approach is to try to describe this relationship through physical models. However, in most cases the underlying physics is too intricate and/or not understood sufficiently to allow the design of feasible physical models which yield satisfying estimation results. In addition, the development of physical models is time consuming, requires precise knowledge about the usually very complex physical processes, and each model is specific for each material and each material transformation process.

To increase cost effectiveness and product quality also of intricate material transformation processes, an approach is necessary which *learns* the underlying physical relationship instead of modeling it based on specific prior knowledge. In addition, it would be highly desirable to have an approach which is generic and can be applied to a variety of materials and transformation processes. In the following we demonstrate that neural networks as adaptive modeling schemes

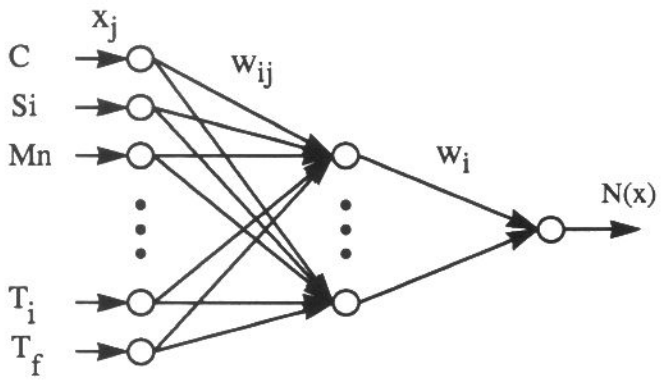

Figure 1: The architecture of the neural network.

have the desired capabilities. We describe the application of a neural network to the problem of estimating the *relative yield stress* (plasticity) of steel plates based on the steel plates's chemical composition, temperature, and shape. Knowledge about the relative yield stress is necessary for optimizing rolling processes, in our case the rolling of steel at a hot line rolling mill.

2. The Neural Network Architecture

The neural network has to model the relation

$$\alpha = F(C, Si, Mn, P, S, Al, N, Cu, Cr, Ni, Sn, V, Mo, Ti, Nb, B, d, b, T_i, T_f)$$

between the relative yield stress α of the steel plate and the concentration of the steel plates's sixteen chemical constituents $C, Si, ..., B$, its thickness d and its width b. T_i and T_f denote the temperature of the steel plate before and after the rolling, respectively. These two temperatures serve as a measure for the actual rolling temperature T, which cannot be determined explicitely. The concentration of the steel plates's sixteen chemical constituents $C, Si, ..., B$ is obtained from a material analysis during the steel cooking.

Figure 1 shows the neural network architecture, a three-layer feedforward network consisting of ten sigmoidal hidden units and one linear output unit. Each hidden unit receives the same twenty-dimensional input vector $\mathbf{x} = (C, Si, ..., Nb, B, d, b, T_i, T_f)$. The weights of the hidden units i, $i = 1, ..., 10$, are denoted by $\mathbf{w}_i = (w_{i1}, ..., w_{i20})$, and the weights of the linear output unit are denoted by $\mathbf{w} = (w_1, ..., w_{10})$. The threshholds of the hidden units and the output unit are denoted by θ_i and θ, respectively. Hence, when the network receives the input $\mathbf{x}$ which carries the information about the steel plate to be rolled, the network generates the output

$$\mathcal{N}_{\mathcal{W}}(\mathbf{x}) = -\theta + \sum_{i=1}^{10} w_i \, \sigma\left(\sum_{j=1}^{20} w_{ij} x_j - \theta_i\right)$$

as an estimation for the relative yield stress of the steel plate, with $\sigma(.) = 1/(\exp(-.) + 1)$ determining the output of the sigmoidal hidden units. The index $\mathcal{W}$ denotes the dependence of the network output $\mathcal{N}_{\mathcal{W}}(\mathbf{x})$ on the set $\mathcal{W} = (\mathbf{w}_i, \theta_i, \mathbf{w}, \theta)$ of all network weights and threshholds.

3. The Adaptation of the Network Weights

The estimation error of the network has to be minimized by selecting the right set of network weights $\mathcal{W} = (\mathbf{w}_i, \theta_i, \mathbf{w}, \theta)$. This is achieved through gradient descent on the mean square

error

$$E(\mathcal{W}) = \frac{1}{2P} \sum_{\mu=1}^{P} (\alpha^{\mu} - \mathcal{N}_{\mathcal{W}}(\mathbf{x}^{\mu}))^2$$

of the last P estimation trials. $\mathbf{x}_{\mu}$ denotes the chemical composition, thickness, width, and temperature of the μ-th steel plate, the actual relative yield stress of which was α^{μ}.

The training data $(\mathbf{x}^{\mu}, \alpha^{\mu})$, $\mu = 1, ..., P$ are accumulated on-line in a FIFO stack (first in first out). With each steel plate which is rolled a new data pair $(\mathbf{x}^{\mu}, \alpha^{\mu})$ is available and put at top of the stack. The data pair at the bottom of the stack, i.e., the data pair which was generated P rolling processes ago, leaves the stack. The stack size P, i.e., the size of the window on the incoming data stream, is determined by the available memeory size and the computation time requirements.

With each update of the FIFO stack by a new data pair the network weights are adjusted through gradient descent on the cost function $E(\mathcal{W})$, which yields, by calculating

$$\Delta \mathcal{W} = -\eta_{dyn} \frac{\partial E(\mathcal{W})}{\partial \mathcal{W}}, \tag{1}$$

the backpropagation learning rules [1, 2]. The step size η_{dyn} is determined dynamically through a line-search procedure.

The need for fast on-line learning was the reason for choosing the stack-like organized training data set together with a line-search procedure for adjusting the weights. Accumulating all available data pairs $(\mathbf{x}^{\mu}, \alpha^{\mu})$ would lead to a training data set which is much too large. About 500 steel plates are rolled per day. Besides the huge memory and computation time requirement, this is also not desirable since the characteristics of the material transformation process might be drifting, and, therefore, very old data pairs might not be representative anymore. On the other hand, to avoid storage of training data completely by performing pattern by pattern learning makes sense only with a very small adaptation step size η_{dyn}. The corresponding slow-down of the adaptation procedure and the increased probability to get stuck in local minima, however, is not acceptable in the described application.

The line-search procedure for determining the optimal value of η_{dyn} looks for the minimum of $E(\mathcal{W})$ along the gradient of $E(\mathcal{W})$ at the current weight set $\mathcal{W}_0$, i.e., it looks for that $\eta_{dyn} = \eta_{opt}$ which minimizes

$$E\left(\mathcal{W}_0 - \eta_{dyn} \frac{\partial E(\mathcal{W})}{\partial \mathcal{W}}\right). \tag{2}$$

Determining η_{opt} based on a quadratic approximation of (2) is sufficient in our case. For that purpose we calculate two support points $(\mathcal{W}_1, E(\mathcal{W}_1))$, $(\mathcal{W}_2, E(\mathcal{W}_2))$ along the gradient in addition to $(\mathcal{W}_0, E(\mathcal{W}_0))$. These three support points define a parabola which forms a quadratic approximation of (2) along the gradient. As the size of adaptation step (1) we choose that $\eta_{dyn} \approx \eta_{opt}$ which minimizes the parabola.

4. The Performance

For testing the performance of the neural network approach and comparing it with the current state-of-the art method, 12000 data pairs $(\mathbf{x}^{\mu}, \alpha^{\mu})$ from the rolling mill were made available by the steel manufacturer. 9000 data pairs formed the training set which was used for adapting the network, and the other 3000 data pairs formed the independent test set. The stack size P was chosen to be 500. The on-line training of the network was simulated by sequentially and randomly picking data pairs $(\mathbf{x}^{\mu}, \alpha^{\mu})$ from the training set, putting the

<E_net>	<E_cur>	Δ	Worst_net	Worst_cur	Δ
34.9%	53.4%	35%	31.7%	36.5%	13.2%

Table 1: The RMS and the worst case estimation error of the neural network and the current state-of-the-art method.

respective data pair at the top of the FIFO stack, and performing an adaptation step (1). Already after having presented about 5000 samples, the network has converged from its randomly chosen initial to its final state. The achieved estimation performance is shown in Table 1. $\langle E_{net} \rangle$ denotes the root mean square (RMS) estimation error of the neural network on the data of the test set, relative to the standard deviation of the test data. $\langle E_{cur} \rangle$ denotes the RMS estimation error of the current state-of-the-art method on the test set, and Δ is the achieved improvement. Table 1 also shows the worst-case, i.e., the largest deviation of the estimation from the real value of the relative yield stress, for the neural network approach as well as for the state-of-the-art method. For the average estimation error the neural network approach yields an improvement of 35%, and for the worst-case the neural network approach yields an improvement of 13.2% over the current state-of-the-art method.

5. Discussion

The results obtained with the neural network approach are very promising. In the application described, the estimation of the relative yield stress of steel, the improvement of the estimation quality is so significant that the neural network approach will replace the currently employed method and soon be a component of a commercially available process optimization system for rolling mills.

There are a couple of reasons for the favorable results with the neural network approach. The main reason is the on-line adaptation of the network. The network weights are permanently adjusted to the changing characteristics of the rolling mill and the drifts of the measuring devices for the chemical composition, thickness, width and temperature of the steel plate. Particularly the calibration of the measuring devices is not very reliable because of the very hazardous environment of a hot line rolling mill.

Another reason for the favorable results seems to be the superiority of feedforward neural networks with sigmoidal hidden units in modeling moderately complex, multivariate functions. To achieve a mean square error of ϵ, a feedforward neural network with $\mathcal{O}(\epsilon^{-1})$ hidden units is sufficient, whereas approximation through series expansion, e.g., polynomial regression, requires at least $\mathcal{O}(\epsilon^{-D/2})$ terms [3]. D denotes the dimension of the input vector $\mathbf{x}$, i.e., the number of input variables, and was 20 for estimating the relative yield stress. D is large in the application domain described, since properties of a material depend on many variables.

References

[1] Werbos P (1974) "Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences." Ph.D. thesis, Harvard Univ. Committee on Applied Mathematics.

[2] Rumelhart DE, Hinton GE, Williams RJ (1986) "Learning Representations by Back-Propagating Errors." Nature, 323:533–536.

[3] Barron AR (1992) "Universal Approximation Bounds for Superpositions of a Sigmoidal Function." IEEE Trans. Information Theory (to appear).