ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 141-146

Neuro-Wavelet Parametric Characterization of Hardness Profiles

V. Colla[†], L.M. Reyneri[††], M. Sgarbi[†††]

[†] Scuola Superiore Sant'Anna, Pisa (I), e.mail: vale@mousebuster.sssup.it
[††] Politecnico di Torino (I), e.mail reyneri@polito.it
[†††] Scuola Superiore Sant'Anna, Pisa (I), e.mail mirkosg@tin.it

Abstract. This work compares a few attempts based on Neural and Wavelet networks, for extracting the Jominy hardness profile of steels directly from the chemical composition. In particular, the paper proposes a multi-networks architecture, where a first network is used as a parametric modeler of the Jominy profile itself, while a second one is used as a parameter estimator from the steel chemical composition.

1. Introduction

Hardenability is one of the main features of steels: it is assessed by means of the *Jominy end-quench test* [1], which produces a vector of hardness values measured along a specimen of heat-treated steel, at predefined positions. These values represent the *Jominy (hardness) profile* and are provided by the manufacturer to characterize steel quality.

Hardenability depends on chemical composition, which is normally analyzed before or during the steel making process. Due to the complex and not yet well understood chemical and physical processes involved, the non-linear relationship between chemical composition and hardness is currently not known.

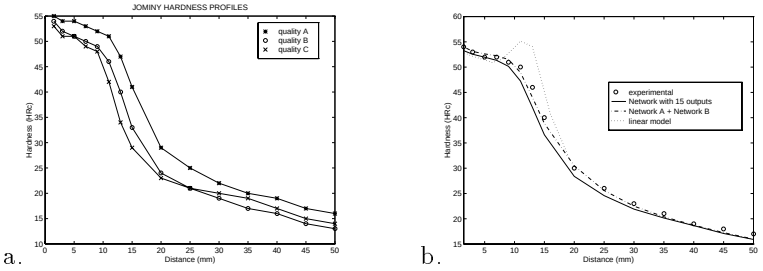
Several models have been developed [2] to predict the shape of Jominy profiles directly from chemical analysis without performing the real test. Most of them are linear, but they are rather inaccurate.

Neural Networks and Wavelet Networks seem to cope well with such a modeling problem. An attempt to apply Neural Networks to predict Jominy profiles has been made in [3] by using a standard Multi-Layer Perceptron with one hidden layer but results were not very good.

This paper presents a more powerful method based on two combined *Neuro-Wavelet Networks* (**NWN**s), where one network provides a parametric model of the Jominy profile, while the second one predicts the parameters for the first one as a function of chemical composition. The extracted parameters do have a strong relationship with the Jominy profile (of which they are a compact representation) and they are also useful to classify steels.

This work has been partially supported by the Italian National Research Council project MADESS II "Architectures and VLSI devices for embedded Neuro-fuzzy control systems"

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 141-146



a. b.

Figure 1: a) A few examples of Jominy hardness profiles for three steel qualities. b) Comparison with estimated Jominy profiles.

2. Jominy Profiles and Preprocessing

Real industrial data were available for three different qualities of Boron steels [2]. Figure 1.a shows a few examples of their typical Jominy profiles (observe resemblance with a sigmoid). Associated with each profile, we have the chemical analysis indicating the content of several micro-alloying elements.

Throughout this work, we consider as input variables the content $\mathcal{C}_i$ of 17 chemical components: C, Mn, Si, P, S, Cr, Ni, V, Mo, Cu, Sn, Al, Ti, B, N, and soluble Al and B. We call:

- $\vec{\mathcal{N}} \triangleq \{\frac{\mathcal{C}_1}{\mathcal{C}_{1,\max}}, \frac{\mathcal{C}_2}{\mathcal{C}_{2,\max}}, \ldots, \frac{\mathcal{C}_{17}}{\mathcal{C}_{17,\max}}, \} \in \Re^{17}$ the vector of *normalized chemical composition*, where $\mathcal{C}_{i,\max}$ is the maximum of $\mathcal{C}_i$ over the training set;
- $J(x)$, the *Jominy profile* as a function of distance x from quenched end.
- $\vec{\mathcal{J}} \in \Re^{15}$, the *Jominy vector* containing the values of $J(x)$ at 15 (sometimes, 18, or 19) predefined positions (often, $x = 1.5, 3, 5, 7, 9$mm, etc.).

To reduce the size of the NWNs used, we tried to reduce as much as possible the number of input variables to the network, without loosing significant information, by means of *Principal Component Analysis* [5]. By retaining the 6 largest eigenvalues, we feed into the NWNs a 6-D vector

$$\vec{\mathcal{V}} = \vec{\mathcal{N}} \cdot \underline{\underline{M}} \tag{1}$$

where $\underline{\underline{M}}$ is a matrix containing as columns the 6 principal eigenvectors.

As a performance index for all NWNs, we adopt the *Normalized Root Mean Square Error* (**NRMSE**) [7]:

$$\epsilon = \sqrt{\frac{\sum_{p=1}^{M}\sum_{j=1}^{N}(\hat{y}_j^p - y_j^p)^2}{\sum_{p=1}^{M}\sum_{j=1}^{N}(y_j^p - \overline{y})^2}} \quad \text{where} \quad \overline{y} = \frac{1}{NM}\sum_{p=1}^{M}\sum_{j=1}^{N} y_j^p \tag{2}$$

where M and N are, respectively, the number of samples in the training (or validation) set and the number of network outputs; y_j^p is the j-th component of the p-th output vector $\vec{Y}^p$ in the training (or validation) set, while $\hat{y}_j^p$ is the corresponding network estimate.

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 141-146

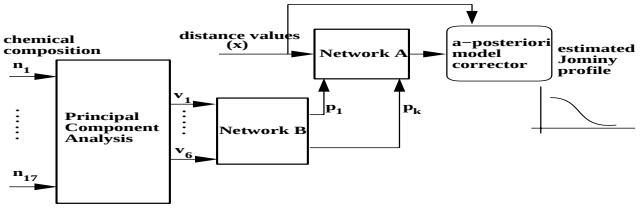


Figure 2: Block diagram of the parametric neuro-wavelet estimator.

3. Parametric Estimation of Jominy Profiles

Let us draw some preliminary considerations on traditional approaches [2, 3]:

- the number of network outputs equals the number of measured Jominy points, namely 15, 18 or 19. But figure 1.a shows that profiles are very smooth. Adjacent values are strongly correlated ($\rho \approx 0.93$), so are also the weight vectors of adjacent neurons (redundant information).
- Approximation errors can produce estimates of the Jominy profiles which are physically not plausible (for instance, small local increases of hardness instead of a monotone decrease).
- The number of points and positions where hardness is measured are not evenly distributed and often differ among different manufacturers, therefore Jominy profiles cannot always be compared directly.
- Hardness measurement $\vec{\mathcal{J}}$ is often affected by large errors.

For all these reasons, we have decided to essay a completely different approach, as sketched in fig. 2. Our system is composed of three interacting blocks:

1. A small 1-in, 1-out NWN (*network A*), used as a *parametric model* of the Jominy profile, which is a function $J(x)$. The set of free parameters of network A (weights, centers and biases) constitutes a vector $\vec{\mathcal{P}}$, which uniquely identifies a Neuro-Wavelet estimate $\hat{J}(x)$ of the Jominy profile $J(x)$ and, consequently, of $\vec{\mathcal{J}}$.
2. An *a-posteriori model corrector* described in section 3.1.
3. A larger NWN (*network B*), used as a *parameter estimator* which predicts the parameter vector $\vec{\mathcal{P}}$ (instead of the profile itself) as a function of chemical composition $\vec{C}$ (namely, vector $\vec{\mathcal{V}}$).

This approach has the following advantages:

- the size of the parameter vector $\vec{\mathcal{P}}$ is much smaller than that of $\vec{\mathcal{J}}$ (see table 1), therefore network B is smaller than would be a network predicting $\vec{\mathcal{J}}$; a smaller training set will suffice, and training and computation times are reduced.

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 141-146

Name	1^{st} layer			2^{nd} layer		size of $\vec{\mathcal{P}}$	NRMSE (%)	
	type	$F(z)$	neu	type	$F(z)$		$\epsilon_{\rm av}$	$\epsilon_{\rm co}$
MLP-1	WRBF-0	hyp. tang.	1	WRBF-0	lin.	4	10.16	4.89
MLP-2	WRBF-0	hyp. tang.	2	WRBF-0	lin.	7	6.76	3.90
LIN†	WRBF-0	hyp.tan.+lin	2	WRBF-0	lin.	5	9.37	4.85
WAV-1	WRBF-2	Mex. hat	1	WRBF-0	lin.	3	16.40	8.53
WAV-2	WRBF-2	Mex. hat	2	WRBF-0	lin.	5	16.08	8.14
RBF-1	WRBF-2	Gaussian†	1	WRBF-0	lin.	3	11.61	6.11
RBF-2	WRBF-2	Gaussian†	2	WRBF-0	lin.	5	9.29	4.97

Table 1: Approximation error (NRMSE) of network A alone, obtained with the proposed NWNs, trained with random initialization and Backpropagation for 1000-5000 epochs. † LIN is a hybrid network (1 lin. plus 1 hyp.tan. neuron in the hidden layer).

- $\vec{\mathcal{P}}$ can be made less sensitive to measurement noise than $\vec{\mathcal{J}}$ (see section 3.1.), therefore steel characterization will be more robust.
- $\vec{\mathcal{P}}$ can also be computed when some measurements of $\vec{\mathcal{J}}$ are missing and it is almost independent of the position of hardness measurements.
- $\vec{\mathcal{P}}$ is as representative of the physical process as $\vec{\mathcal{J}}$, therefore it can be used to classify (more robustly) steel quality.

3.1. Choice of the Parametric Model (Network A)

The choice of the best NWN for network A (parametric model) is by itself not a simple problem, due to the need of reducing as much as possible the number of tunable parameters while maintaining a good estimation and classification accuracy. We have therefore essayed a set of very small two-layers WRBFs [4].

Table 1 (column $\epsilon_{\rm av}$) compares the different models analyzed. The values given are an average over the whole data set (615 different specimen).

By looking at the outputs of networks A in details, we have observed that the estimation error of each element of the Jominy profile predicted by network A does not have a null average, and the average itself varies with distance x. We have therefore decided to subtract this average modelization error (in tabular form) from the output of network A (*a-posteriori correction*), as shown in fig. 2. This has reduced the modelization error roughly by a factor 2, as indicated in table 1 (column $\epsilon_{\rm co}$).

We have then selected two networks (namely, MLP-1 and LIN), according to the following criteria:

1. to have the smallest approximation error. Networks MLP-1, MLP-2, LIN and RBF-2 are the best under this respect, thanks to the particular shape (nearly sigmoidal) of the Jominy profile.
2. To have as few parameters as possible. This reduces both training time for network A and the size of network B. Networks WAV-1, RBF-1, MLP-1 are the best under this respect.

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 141-146

3. To have a set of parameters which are as representative as possible of the physical process of hardening. The degree of representativity has been assessed by analyzing the correlation between pairs of Jominy vectors $\vec{\mathcal{J}}$ and the corresponding parameter vectors $\vec{\mathcal{P}}$. The closer two profiles $\vec{\mathcal{J}}$ are to each other, the closer parameters $\vec{\mathcal{P}}$ should be. Networks MLP-1 and LIN networks are the best under this respect (results not shown here).

4. To have a set of parameters which provide the smallest *noise sensitivity*, namely the smallest sensitivity to the noise affecting hardness measurements. That is assessed by measuring the effects of random noise, superimposed to the measured data, on the parameters and consequently on the reconstructed profile. Simulations (not shown here) proved that noise on the reconstructed profile is less than 50% of measurement noise. Networks LIN, WAV-1, MLP-1 and EXP-1 are better under this respect.

4. Parameter Estimation (Network B)

¿From the 6 principal components $\vec{\mathcal{V}}$ of chemical composition $\vec{\mathcal{N}}$, network B (see fig. 2) should predict a set of parameters $\vec{\mathcal{P}}$ characterizing the corresponding Jominy profile. The issue of choosing the most suitable network to perform this association is a problem of approximating an unknown non-linear function $\Re^6 \rightarrow \Re^4$, or $\Re^6 \rightarrow \Re^5$, respectively for an MLP-1 and a LIN network A.

We have tested different kinds of two-layer networks, namely Wavelet Networks (**WAV**s), Multi-Layer Perceptron (**MLP**s) and Radial Basis Functions (**RBF**s) with Mexican hat, hyperbolic tangent and Gaussian activation functions, respectively [4], and with different number of neurons in the hidden layer. Initialization is either random (for MLPs) or OLS [6] (for WAVs and RBFs).

Training and validation sets contain 615 and 152 samples, respectively, distributed over the three considered qualities. As the work is still going on, we are collecting more data that will be inserted only in the validation set.

Table 2 shows the results on the validation set for different types of network B. For both MLP-1 and LIN networks A, the table gives: the NRMSE on parameter estimation ϵ_{par}, which depends on network B alone; the NRMSE on Jominy vector estimation, with (ϵ_{co}) and without (ϵ_{av}) the a-priori model correction shown in fig. 2. The latter two depend on both networks (A and B). The last column of the table gives, as a comparison, the NRMSE obtained with only one network (with 15 outputs) predicting the Jominy vector directly as a function of chemical composition (traditional approach).

WAV networks with 4 wavelons, associated with the corrected LIN network A gives the overall best performance. Comparable results can be achieved with RBF+LIN and WAV+MLP-1 networks. MLP are the worst among all.

Also the unique network (traditional approach, last column) gives comparable results. This proves that our approach gives performance comparable with other existing methods, but offers the several advantages outlined in section 3. which are very worthy in real industrial applications.

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 141-146

Network B	hid	MLP-1			LIN			1 net
	neu	$\epsilon_{\mathrm{par}}(\%)$	$\epsilon_{\mathrm{av}}(\%)$	$\epsilon_{\mathrm{co}}(\%)$	$\epsilon_{\mathrm{par}}(\%)$	$\epsilon_{\mathrm{av}}(\%)$	$\epsilon_{\mathrm{co}}(\%)$	$\epsilon(\%)$
WAV	2	3.66	18.4	15.2	3.42	18.20	15.8	
WAV	4	3.53	18.3	15.1	3.49	16.39	13.5	13.9
WAV	8	3.45	17.5	13.7	3.49	16.24	13.6	13.4
WAV	16	3.59	16.8	13.6	3.59	16.46	13.8	13.7
RBF	2	3.43	18.5	15.9	3.40	18.3	16.0	
RBF	4	3.48	18.5	15.6	3.52	16.7	14.5	14.1
RBF	8	3.42	18.7	16.2	3.43	16.3	14.0	14.0
RBF	16	3.50	17.0	13.9	3.47	16.5	14.0	14.1
MLP	2	3.46	19.6	17.1	3.37	19.0	17.0	
MLP	4	3.53	19.4	16.8	3.40	42.0	36.5	14.3
MLP	8	3.46	71.6	66.8	3.42	47.5	43.6	13.8
MLP	16	3.36	39.5	37.6	3.99	26.5	25.3	15.0
linear model [2]	$\epsilon = 39.4\%$							

Table 2: Overall performance (NRMSE) on the validation set. Last column gives results obtained with traditional neural methods [3].

Table 2 (last row) also compares our results with those achieved by the linear model for Boron steels proposed in [2], which provides a much higher NRMSE than our method. Furthermore, the validity of that linear model is guaranteed only when the concentration of certain chemical components is within very restrictive ranges.

Figure 1.b compares a measured Jominy profiles with the corresponding prediction, for the proposed and the linear methods. This confirms that, in practical cases, the neuro-wavelet parametric approach can be a more effective and reliable alternative to traditional models.

References

[1] "Standard Method for End-Quench Test for Hardenability of Steel," A255, *Annual Book of ASTM Standards*, pp. 27-44, ASTM, 1989.

[2] D.V. Doane J.S. Kilkaldy (eds.): "Hardenability Concepts with Applications to Steel", TMS-AIME, 1978.

[3] W.G. Vermeulen, P.J. van der Wolk, A.P. de Weijer, S. van der Zwaag: "Prediction of Jominy Hardness Profiles of Steels Using Artificial Neural Networks," *Journal of Materials Engineering and Performance*, Vol. 5, No. 1, February 1996.

[4] L.M. Reyneri, "Unification of Neural and Wavelet Networks and Fuzzy Systems", submitted to *IEEE Trans. on Neural Networks*, 1998.

[5] W.W. Cooley, P.R. Lohnes: "Multivariate Data Analysis," John Wiley & Sons Inc., USA, 1971.

[6] S. Chen, C.F.N. Cowan, P.M. Grant: "Orthogonal Least Squares Learning Algorithm for Radial Basis Function Networks," *IEEE Transactions on Neural Networks*, Vol. 2, No. 2, pp. 302-309, March 1991.

[7] Q. Zhang, A. Benveniste: "Wavelet Networks," *IEEE Transactions on Neural Networks*, Vol. 3, No. 6, pp. 889-898, November 1992.