PERIODICA POLYTECHNICA SER. EL. ENG. VOL. 47, NO. 1–2, PP. 141–147 (2003)

SIMPLIFYING THE MODEL OF A COMPLEX INDUSTRIAL PROCESS USING INPUT VARIABLE SELECTION

Nóra SZÉKELY

Department of Measurement and Information Systems
Budapest University of Technology and Economics
H–1521 Budapest, Hungary, P. O. Box 91.
szekelyn@mit.bme.hu

Received: September 15, 2002; Revised: February 25, 2003

Abstract

This paper deals with experience gained from building a neural model of a Linz-Donawitz (LD) steel converter. The complexity of the process makes this task difficult because many variables affect the quality of the resulted steel. The paper details the simplification of the neural model using input variable selection (IVS) methods. Three types of models were investigated: one using the originally measured physical parameters, and two types using transformations, namely independent component analysis and principal component analysis. Transformations were applied to derive new parameter spaces where the importance of parameters shows higher differences. The relevance of the original and the transformed parameters were measured by different ways.

Keywords: neural modelling, input variable selection.

1. Introduction

Neural networks are one of the possible means to build complex nonlinear mappings between many inputs and some outputs. Experimental data are used to train the network, until its operation will be similar to that of the real industrial process. One of the most important steps of building a model for an industrial problem is to construct reliable database that is to select and preprocess the experimental data. These steps are very important if the data contain noisy, imprecise information and where the problem space is rather large.

The industrial modelling problem we dealt with was a steel production process with an LD converter. Steel production is a complex process where many variables affect the quality of the result. There are many input parameters registered and two essential output parameters: the carbon content of the steel and its temperature at the end of the blasting process [4].

The quality of steel is mainly determined by the amount of oxygen used during blasting. The acceptable ranges of the output parameters are narrow, so it is an important and hard task to create a reliable predictor to determine the amount of oxygen needed to obtain predetermined quality. To give a reliable prediction, we have to know the relation between the input and output parameters of the process, so we have to build a model of steel production. The output parameter of primer importance is the final temperature of steel, thus, the developed model has only one output. Models using all input data are very complex while some of the inputs may be irrelevant so the number of parameters is to be decreased. The input data can be applied in their original form or some transformation can be performed and the resulting parameters can be used.

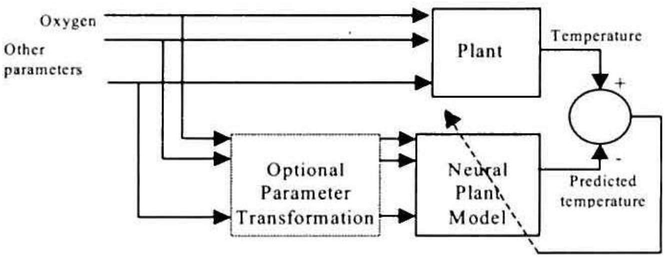

Fig. 1. The temperature (forward) model of the LD steel production process

This task is difficult because many effects cannot be taken into consideration exactly, therefore conventional methods (mathematical models based on physical and chemical laws), or even expert systems fail. What is only known is that there is a nonlinear relation between the input and the output, so a neural model seems to be appropriate. The model should be as simple as possible, therefore the irrelevant input parameters are not to be used. Basically three models were investigated:

- One using the original physical parameters,
- One based on principal component analysis (PCA),
- One based on independent component analysis (ICA).

2. Parameter Transformation Techniques used: PCA and ICA

In this section a brief description of PCA and ICA methods is given. Both techniques use linear transformation to produce a new data set, but the resulted parameters are believed to perform better than the original ones.

The main goal of PCA (principal component analysis) is to provide a new parameter space in which the dimension-reduction of the data is much easier than in the original space. The basic idea is that parameters having high variance carry the lion's share of information contained in the data, while parameters which are close to be constant are less important therefore can be omitted. In many cases dimension reduction is very hard (or impossible), because it is difficult to find parameters that have little variance. PCA is a transformation that results in such an orthogonal basis that solves this problem. The new basis is found by diagonalizing the centered covariance matrix (C) of the original data set. The new coordinates in the eigenvector basis are called *principal components*. The size of an eigenvalue

λ corresponding to an eigenvector v equals the amount of variance in the direction of v. Furthermore, the directions of the first n eigenvectors corresponding to the first n biggest eigenvalues of C cover as much variance as possible by n orthogonal directions [1]. This technique can help to simplify the model, because it is assumed that parameters with little variance generally carry irrelevant information, and the very parameters are omitted in the model.

The ICA (independent component analysis) method also produces linear combination of the original parameters, but the resulted parameters have a different property, namely, they are *statistically independent* [2]. It is easy to see why we are interested in independence in the case of modeling an industrial process: we look for the independent effects that influence the output. Since some of the originally measured parameters depend on each other (for example, the amount and temperature of pig iron, and the amount of the necessary oxygen), it is believed that with the help of ICA, the number of necessary inputs can be significantly reduced. However, to obtain these components, independence should be somehow measured. It is hard, so instead of independence, nongaussianity is measured. The reason why this is allowed lies in the central limit theorem. It says that the sum of independent random variables tend towards a Gaussian distribution. Thus, the sum of two independent random variables usually has a distribution that is closer to Gaussian than any of the two original ones. Therefore, finding maximally non-Gaussian components, in most cases, equals to finding maximally independent ones. Several techniques can be used, like the classic method based on kurtosis, or other methods based on entropy of the variables [2].

The definition of kurtosis:

$$\mathrm{kurt}\,(y) = E\{y^4\} - 3(E\{y^2\})^2. \tag{1}$$

Kurtosis is zero for Gaussian random variables, and nonzero for most non-Gaussian variables, so it can be used to measure independence. This is the classic method, but it can be very sensitive to erroneous or irrelevant observations, which are frequent in the case of industrial processes, so a better technique has to be found.

Techniques based on the measurement of entropy (H) are more robust. A fundamental result of information theory is that a Gaussian variable has the largest entropy among all random variables with equal variance. This means that entropy can be used as a measure of nongaussianity. In most cases, a slightly modified form of entropy is used, called negentropy (J). It is zero for Gaussian variables and always positive for non-Gaussian variables, so it is a technically better measure. Negentropy is defined as:

$$J(y) = H(y_{\mathrm{gauss}}) - H(y). \tag{2}$$

Here y_{gauss} is a Gaussian random variable with the same covariance as y. Calculating negentropy is difficult, but good approximations can be found, for example:

$$J(y) \propto [E\{G(y)\} - E\{G(\nu)\}]^2, \tag{3}$$

where G is a suitable non-quadratic function, and ν is a Gaussian random variable with zero mean and unit variance.

The implementation used in this work was the FastICA algorithm [2].

3. Measuring the Relevance of Input Parameters

Many input parameters are measured, and it is a really hard task to decide which of them have a significant influence on the output. Input variable selection (IVS) is a method that can help to determine which parameters are required for building a reliable and sufficiently good model using the possible lowest number of parameters.

There are many approaches to select important variables, three were applied here:

- Input parameter selection using the expertise of steel production experts.
- Measuring the importance of input parameters by calculating the cross-correlations between the input parameters and the output.
- Measuring the importance of input parameters by calculating fourth-order cross-cumulants between the input parameters and the output.

Selection of measured physical variables offered by experts can be carried out only on the original database. The problem with this approach is that there is no exact explanation why the very parameters should be used, and it turned out that the importance of some parameters were over- or underestimated by the experts.

Using cross-correlation coefficients (R) is very simple, and it seems to be rather effective. Here relevance is measured by the covariance of the standardized input and output variables:

$$R(x, y) = \frac{\operatorname{cov}(x, y)}{\sigma_x \cdot \sigma_y}. \tag{4}$$

Another method applied to find relevant parameters is based on fourth-order cross-cumulants (suggested by A. D. BACK and A. CICHOCKI), that can be defined as the following [3]:

$$C_{xyxx} = E\{xyxx\} - 3E\{xx\}E\{xy\}. \tag{5}$$

It is important that both input x and output y should be normalized zero-mean variables, otherwise the test fails.

4. Results

A data set of about 4500 steel production charges was used to check the results of input variable selection techniques. The data set contained 52 parameters for each record; all the physically interpretable 31 input ones and one output parameter (the desired temperature) were selected for use.

The neural network used was a simple feedforward MLP, with 10 input neurons and 5 hidden neurons. The learning method was a slightly modified backpropagation (BP with moment). This learning method can be classified as a strongly simplified conjugate gradient method. Previously, several NN topologies and architectures (MLPs of various size, RBFs) were examined with several optimization methods, but considering performance, none of them proved to be significantly better than the others. So the accuracy of the model seems to be limited mainly by the available data set. With respect to the speed of learning, the algorithms definitely show differences, but in this special case, the accuracy of the model was of primary importance, therefore we decided to use the very simple BP with moments learning method.

Three types of models were investigated, one using the original physically interpretable parameters, and two model types with some transformation. In one transform type model, PCA was applied, in the other, ICA was used.

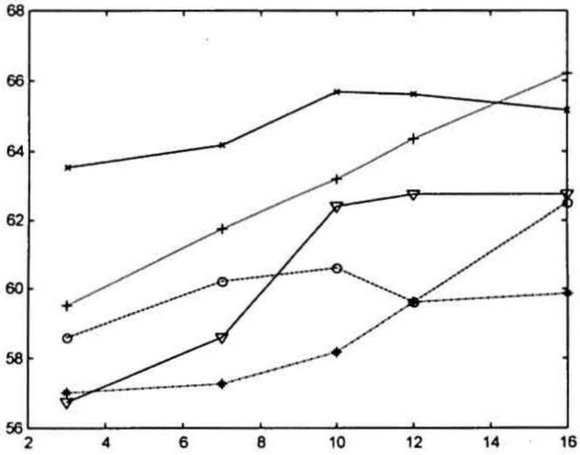

Fig. 2. Performance of different models in the case of 3, 7, 10, 12 and 16 input parameters

The important input parameters were selected first, then the model using that input parameters was trained. 75% of the data set was used as training examples; the remaining 25% was applied as test set. The generalization error measured on the test set characterized the usefulness of the input selection method.

In the horizontal axis of *Fig. 2*, the number of input parameters (original or transformed) used is shown, while on the vertical axis, the measure of model correctness (measured by the percentage of appropriate temperature predictions on the test set).

The selection of original parameters was based on experts' knowledge. In *Fig. 2* the data points denoted by 'o' and a dashed line correspond to the performance of such models. Five such models of 3, 7, 10, 12 and 16 input parameters were tested. For comparison, the correctness of the models using all the available 31 input parameters is about 66%.

The selection of PCA parameters was based on the cross-correlation of the PCA components and the output temperature. In *Fig. 2* the data points denoted by '+' and a dotted line corresponds to the results of such models. It should be emphasized that the common idea of using the PCA components corresponding to the highest eigenvalues performed much worse. (It is not true, in this case, that components of highest variances have the highest useful information content.) In *Fig. 2* the data points denoted by a triangle and a solid line show the performance of these parameters. The PCA parameters cannot be selected based on the cumulant measure, because the necessary conditions are not met.

Two model types of utilizing the same ICA transform were investigated; the difference is in the input variable selection method. One is based on the cross-correlation of the ICA parameters and the output. In *Fig. 2* the data points denoted by 'x' and a solid line corresponds to the results of such models. The other selection method was based on the cumulant measure. In *Fig. 2* the data points denoted by '*' and a dashdot line corresponds to such models.

Both PCA and ICA transform parameters selected by using cross-correlation perform better than the physical parameter model, especially in the case of small number of inputs. Using only 10% of the available parameters (3 out of 31) the most correlated independent components (ICA) give nearly perfect performance compared to models of 31 parameters. The cross-cumulant technique (applied on ICA components) works quite badly, it gave the worst results in almost all cases. Test results are shown also in *Table 1*.

Table 1. Performance of different models in case of 3, 7, 10, 12 and 16 input parameters

	Performance	Number of inputs				
	(in %)	3	7	10	12	16
Model type	Original	58.62	60.23	60.60	59.61	62.47
	PCACorr	59.52	61.75	63.18	64.34	66.22
	PCAMax	56.75	58.62	62.38	62.73	62.73
	ICACorr	63.54	64.18	65.68	65.59	65.15
	ICACum	57.02	57.28	58.18	59.61	59.87

5. Conclusion

This paper details the simplification of a neural model using input variable selection (IVS) methods. Three types of models were investigated: one using the originally measured physical parameters, and two types using transformations, independent component analysis and principal component analysis. The relevance of the original and the transformed parameters were estimated by three methods: utilizing the human expertise about the process, using cross-correlation technique and using cross-cumulant technique.

The results of the work show that ICA components give the best results in the case of small number of inputs. Both PCA and ICA based models performed better than models using the original parameters. Against the common idea, PCA parameters should be selected applying cross-correlation, the high eigenvalue does not guarantee relevant information content.

References

[1] MIKA, S. – SCHÖLKOPF, B. – SMOLA, A. – MÜLLER, K.-R. – SCHOLZ, M. – RÄTSCH, G., Kernel PCA and De-noising in Feature Spaces. In: Kearns, M. S., Solla, S. A., Cohn, D. A., (Eds.), *Advances in Neural Information Processing Systems* 11, Cambridge, MA, pp. 536–542. MIT Press, 1999.

[2] HYVÄRINEN, A. – OJA, E., Independent Component Analysis: Algorithms and Applications. *Neural Networks*, **13** (4–5) (2000), pp. 411–430.

[3] BACK, A. D. – CICHOCKI, A., Input Variable Selection Using Independent Component Analysis and Higher Order Statistics, *Proc. of the First International Workshop on Independent Component Analysis and Signal Separation – ICA'99* , Aussois, France, Janurary 11–15, 1999, pp. 203–208.

[4] STRAUSZ, GY. – HORVÁTH, G. – PATAKI, B., Effects of Database Characteristics on the Neural Modelling of an Industrial Process, *Proc. of the International ICSC/IFAC Symposium on Neural Computation/NC'98* , Vienna, Sept. 1998, pp. 834–840.