Analytica Chimica Acta, 277 (1993) 289–295
Elsevier Science Publishers B.V., Amsterdam

Artificial neural networks as a multivariate calibration tool: modeling the Fe–Cr–Ni system in x-ray fluorescence spectroscopy

A. Bos, M. Bos and W.E. van der Linden

Department of Chemical Analysis, University of Technology Twente, P.O. Box 217, 7500 AE Enschede (Netherlands)

(Received 3rd September 1992)

Abstract

The performance of artificial neural networks (ANNs) for modeling the Cr–Ni–Fe system in quantitative x-ray fluorescence spectroscopy was compared with the classical Rasberry–Heinrich model and a previously published method applying the linear learning machine in combination with singular value decomposition. Apart from determining if ANNs were capable of modeling the desired non-linear relationships, also the effects of using non-ideal and noisy data were studied. For this goal, more than a hundred steel samples with large variations in composition were measured at their primary and secondary K_α and K_β lines. The optimal calibration parameters for the Rasberry–Heinrich model were found from this dataset by use of a genetic algorithm. ANNs were found to be robust and to perform generally better than the other two methods in calibrating over large ranges.

Keywords: Multivariate calibration; Principal component analysis; X-ray fluorescence spectrometry; Artificial neural networks; Chromium; Iron; Neural networks; Nickel; Steel

Artificial neural networks (ANNs) have gained much focus in recent years due to their ability to 'learn' arbitrary non-linear relationships between input and output spaces. Applications of ANNs in current literature can be divided in two distinct classes: as pattern classification systems, and as 'function approximation' systems. The latter is commonly referred to in the field of chemometrics as multivariate calibration.

The application of instrumental techniques in analytical chemistry usually implies the need for calibration in order to obtain accurate results. When the relationship (or model) between measured signals and concentrations is known to be linear, or can be made linear through transformation, standard methods can be applied for calibration of which principal components regression analysis (PCR) and singular value decomposition (SVD) are best known. In other cases a non-linear model must first be determined of which the characteristic parameters can be found through appropriate calibration methods. The simultaneous quantitative determination over large ranges of iron, nickel and chromium concentrations in stainless-steel samples with x-ray fluorescence spectroscopy is such a case. Due to strong interelement effects between these components the concentrations are not proportional to their relative signals and a non-linear model is required.

In this paper, three empirical approaches to solve this problem are compared: the approach proposed in the paper by Rasberry and Heinrich

Correspondence to: W.E. van der Linden, Department of Chemical Analysis, University of Technology Twente, P.O. Box 217, 7500 AE Enschede (Netherlands).

[1] which is based on empirically modeled interelement effects, a linear-based approach using singular value decomposition on the set of signals and their cross-products, and an approach in which the measured signals are used directly as inputs to a neural network which provides the desired concentrations at its outputs.

THEORY

Measured intensities in x-ray fluorescence analysis are usually not proportional to the concentrations in the specimen. Apart from inhomogenous and particulate samples, this can occur when the x-ray emission of the analyte is significantly affected by concentration variations of the other elements in the sample. These interelement effects depend on the mass fractions of the elements in the sample, and the relationships of their absorption coefficients for the primary and secondary radiation. Also secondary fluorescence can occur. These effects are illustrated in Fig. 1. In order to obtain accurate results, a non-linear calibration model is required. These models can be divided in two categories, one based on equations derived from theoretical principles and physical constants, and the other based on empirical calibration using standards. In the Cr–Ni–Fe system found in most alloys, strong interelement effects occur. The FeK_α is absorbed by chromium, and secondary fluorescence is induced in iron by the NiK_α. The NiK_α is absorbed in turn by chromium and iron, whereas iron and nickel both induce secondary fluorescence in chromium. In the following sections, three approaches for modeling this complex system are explained and compared.

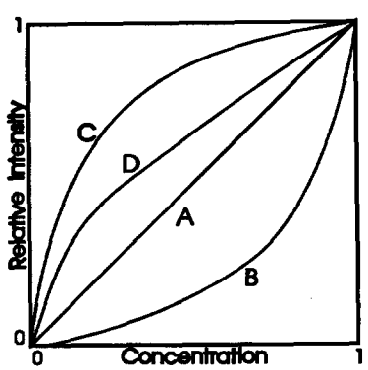

Fig. 1. Types of interelement effects in x-ray fluorescence spectroscopy: (A) linear calibration curve; (B) preferential absorption by matrix; (C) preferential absorption by analyte; (D) secondary fluorescence.

The Rasberry–Heinrich model

In this model the following empirically derived relationship is proposed

$$\frac{C_i}{R_i} = 1 + \sum_{k \neq i} A_{ik} \cdot C_k + \sum_{k \neq i} \frac{B_{ik} \cdot C_k}{1 + C_i} \qquad (1)$$

in which C_i and C_k denote the concentrations (fractions) of components i and k, R_i the relative intensity of component i, i.e., the measured intensity divided by the intensity found for the pure element, and A_{ik} and B_{ik} are calibration coefficients compensating for absorption and secondary fluorescence effects respectively. In their paper, the authors apply only the coefficients of the *dominant* effect, thus leading to *six* coefficients in a ternary system instead of the twelve possible. The calibration coefficients have to be determined empirically from a set of samples with known compositions. After calibration, unknown concentrations can be calculated by setting the initial concentrations equal to the relative intensities found and then iteratively recalculating new concentrations through Eqn. 1 until convergence is achieved. A flow chart of the algorithm can be found in [1]. The calibration can either be done by solving the set of equations or by graphical methods [1], but this would require too much effort when using a leave one out approach. Therefore, in this work a genetic algorithm (GA) is applied as combinatorial optimization technique for finding the required calibration constants. A full description of GAs can be found in [2,3], and an application to analytical chemistry in [4]. Here only a short description of the characteristics is given.

A GA is a stochastic combinatorial optimization method based on an artificial simulation of the Darwinistic evolutionary process which pro-

vides a robust and efficient method for searching very large parameter spaces. Randomly initialized sets of parameters are mapped onto 'chromosomes', and an evolutionary process is started in which all individuals (sets of parameters) compete for reproduction through a 'fitness function'. By using crossover functions, new generations are capable of combining beneficial subparts of the chromosomes to create better fitting individuals, thus 'evolving' to an optimal solution. Mutation is applied to keep diversity in the gene pool. Whereas the choice of a correct cooling curve is essential in the case of the comparable 'simulated annealing' method, the choice of pool size, fitness function and crossover and mutation probabilities is essential for GA's.

A semi-linear approach

As mentioned before, the concentrations are not directly proportional to the measured intensities, thus prohibiting the application of linear calibration. This can be overcome by expanding the data set with quadratic and cross product terms of the measured intensities as has been demonstrated earlier [4]. After measuring a set of samples, a relationship can be made for each element of interest as in Eqn. (2).

$$\mathbf{A} \cdot \vec{x} = \vec{c} \tag{2}$$

The matrix **A** contains in each row the measured intensities and their quadratic and cross product terms, and the vector $\vec{c}$ in the same row the corresponding concentration. An optimal solution vector $\vec{x}$ can then be determined through means of singular value decomposition [5].

Neural networks

Artificial neural network theory supplies us with a general adaptive model for *learning* an arbitrary non-linear mapping from an input space to an output space. This is done by simulating a network topology and then presenting it with a database of *examples* and applying a learning rule. Through the learning rule, the network will adapt and learn from the examples to respond correctly to its environment. Especially the perceptron-based backward error propagation networks, or backpropagation networks, are widely used in many areas and as such are the focus of intensive research and numerous papers. A short recapitulation of this type of network and its properties will be given here. Descriptions of backpropagation networks applied to analytical chemistry have been given in earlier papers [6,7]. For a more explicit description refer to Ref. 8.

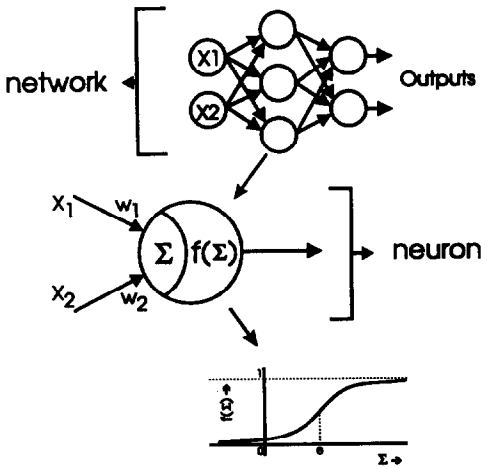

Fig. 2. A typical backpropagation network consisting of sigmoid neurons.

Backpropagation networks are mainly characterized by the type of units (neurons) used, by the topology [interconnections and their weights (or strengths)] of the network, and by the learning rule applied. The basic units, of which the 'sigmoid neuron' is best known, perform a weighted processing of the input signals, which are either pattern inputs or outputs of other neurons. This is followed by a non-linear scaling or squashing function as is illustrated in Fig. 2 and Eqns. 3a–3b.

$$net = \sum_i x_i \cdot w_i \tag{3a}$$

$$f(net) = \frac{1}{1 + e^{-net+\theta}} \tag{3b}$$

The biases (θ_i) for each neuron are usually replaced by extra weights to an imaginary constant input of magnitude 1 for easier programming. A network is built by combining these units into layers, and then applying at least one layer, a

so-called 'hidden' layer, of neurons between the input and the output layer of the network. The non-linearity of the squashing function and the occurrence of at least one hidden layer are required to allow a network to learn a non-linear mapping. The free variables in a network are the weights, or interconnection strengths, usually represented as floating point numbers. The goal of the learning rule is to find a set of weights for a given network topology by initializing the weights to random values and then repeatedly adapting them until the network produces the correct outputs for each set of inputs in the examples. In general there are many, if not unlimited, possible sets of weights leading to the same global behavior of the network. Which one is found depends on the random initialization of the weights before the training starts and the implementation of the training phase.

The iterative process of adapting the weights is usually a time-consuming task. The adaptation of the weights in the original backpropagation algorithm is done by first order gradient descent, with a fixed step size parameter chosen by the experimenter, which converges very slowly near the optimum due to the small gradient. This effect can be countered by introducing an ad hoc momentum factor. To obtain accurate *quantitative* results, one must continue training for very long times. In recent years, several adaptations of the conjugate gradient methods, which apply second order information in determining step directions and sizes, have been proposed for backpropagation networks [9,10]. Of these, the 'scaled conjugate gradient' (SCG) method as proposed by Möller [10] has been implemented by us with good results. Möller's proposal applies a Levenberg–Marquardt approach in scaling the step size as opposed to the more common, and computationally intensive, line search found in other papers. The resulting algorithm, although rather complex, requires no external or empirical parameters to be set like the learning rate and momentum in the original backpropagation. The high rate of convergence, theoretically quadratic near the optimum for conjugate gradient methods, enables achieving accurate quantitative results in acceptable calculation times.

EXPERIMENTAL

The data set

About 130 stainless-steel samples of which the composition varied widely were provided to us by the faculty of Mechanical Engineering at our university. Apart from iron, nickel and chromium, also relatively large amounts of manganese, cobalt and vanadium were present, as well as traces of many other elements. Of the samples, a varying number of components had been determined in the past through wet-chemical analysis with unknown accuracy. All samples had polished surfaces and were measured quantitatively in a Philips PW1480 spectrometer under control of the Philips X44 software running on a PDP-11 minicomputer. The x-ray tube had a chromium anode, therefore an aluminum filter of 0.085 mm thickness was used when measuring near the chromium fluorescence lines to compensate for the chromium radiation from the tube itself. For all elements an LiF-200 crystal and fine collimator were used. Nickel and the primary K_α of chromium were detected by flow counter and scintillator, and the others by flow counter alone. To measure the fluorescence of the elements with low concentrations sufficiently, the tube was operated at 50 kV and a current of 50 mA. All primary and secondary K_α and K_β lines of all major components were measured, although only parts of these data were used during the comparison tests. Also intensities of pure samples of all elements found in the steel samples were determined. When possible directly from the pure element itself, in other cases through compound standards and calculations. The high operating current of the tube saturated the detector when determining the strong fluorescence of the primary lines of pure iron and nickel, therefore for these elements a linear calibration at lower currents was done and an extrapolation to 50 mA made.

Of the acquired data, those samples with obvious errors such as total compositions of much more than 100% and those of which the rest (the components which are not iron, nickel or chromium) exceeded 5% of the total were taken out. The remaining 101 samples contained com-

TABLE 1

Concentrations found in the steel samples

	Min. (%)	Max. (%)
Iron	50.48	98.17
Nickel	0.02	21.80
Chromium	0.04	29.20
Rest	0.72	4.03

positions varying over a large range, as illustrated in Table 1.

Procedures and parameters

All models were tested using a 'leave one out' approach to gain maximum information from the available data. In this approach one pattern is removed from the data set and the model is calibrated with the remaining patterns. The pattern taken out is then evaluated as an 'unknown' and the errors are calculated. Then the pattern is put back and the whole process is started anew for the next one in the data set. In this manner each pattern serves as an unknown as well as a calibration pattern.

For the Rasberry–Heinrich model, the Cr$K_{\alpha 1}$, the Ni$K_{\alpha 1}$, the Fe$K_{\alpha 1}$ and respective concentrations were selected from the original data set. Two leave one out experiments were done, one as a Cr–Ni–Fe ternary system, and in the other the remaining components were treated as an imaginary fourth component 'X' as suggested in the original paper [1]. In the former case, the six coefficients as suggested by the authors, and in the latter case six extra coefficients for absorption and secondary fluorescence effects of X were determined. This was done because no dominant effect of X was known. In the algorithm, the fraction of X was continuously calculated as one minus the fractions of Cr, Ni and Fe. If the fraction of X became negative, it was set to zero. As a check for the application of the genetic algorithm in finding the calibration constants, a test run was performed with the data as published in the original article [1].

To decrease the calculations to manageable proportions for the genetic algorithm, first an approximation of the calibration coefficients was derived from the whole data set using a population of 60 individuals with 16 gray-coded bits allocated for each coefficient scaled in the interval [−3.0, 3.0]. The algorithm was allowed to continue until no more improvement was observed. Some random noise was added to the coefficients found and they were then used as a base in the leave one out method. In this method, a population size of 40 was used, and offsets from the previously found base-coefficients were gray-coded as 8 bits scaled in the interval [−0.5, 0.5], resulting in a resolution of approximately 0.004 for each coefficient. The crossover rate was set to a chance of 0.2, and the mutation rate to a probability of 0.01 per bit location. After some testing, the number of generations was fixed at 50, allowing full convergence for all samples in the data set.

For the semi-linear test, the same signals as for the Rasberry–Heinrich model were used expanded with all possible quadratic and cross product terms. Singular value decomposition was used with a varying number of terms in the diagonal matrix to find optimal solution vectors $\vec{x}$ for each given set of training examples. Also the inclusion of intensities for the trace elements (Mn, Co and V) were tried, but this led to slightly worse results than with the intensities for the three main elements alone.

All neural nets were of the fully connected feed forward type employing the sigmoid function as squashing function. Primary K_α and K_β of Cr, Ni, Fe, Mn, Co and V were used as input signals. All inputs were scaled in the interval [0.1, 1.0] and the outputs were scaled in the interval [0.1, 0.9]. Scaling of the outputs is required as the sigmoid function is asymptotic near 0 and 1. The scaling of the inputs was not required in this case, especially as the signals were relative, but is an automatic feature of our software. Preliminary experiments, however, showed no increase in performance when using unscaled signals for training. All networks were trained for 15 000 epochs using the scaled conjugate gradient algorithm allowing maximum convergence. Different networks were used for each of the three components to simplify optimization and a range of network topologies were tested. The best results were used for the comparison.

TABLE 2

Comparison of calibration constants found by the genetic algorithm (GA) for the dataset and those as published by Rasberry and Heinrich [1]
(A denotes absorption and B denotes secondary fluorescence effects)

Effect	Type	Coefficients	
		[1]	GA
Cr → FeK_α	A	2.10	2.11
Cr → NiK_α	A	1.20	1.21
Fe → NiK_α	A	1.71	1.75
Fe → CrK_α	B	−0.46	−0.40
Ni → FeK_α	B	−0.47	−0.44
Ni → CrK_α	B	−0.27	−0.23

RESULTS AND DISCUSSION

In Table 2 the calibration constants found by the genetic algorithm for the dataset of binary and ternary samples as originally published by Rasberry and Heinrich [1] are given, as well as the constants they found. Several runs with different random starting populations converged invariably to the same set of coefficients. The mean error for the set of five unknowns are compared in Table 3. Although with the GA slightly different constants are found, the results in predicting the unknown samples are better.

With the dataset of steel samples, however, a distribution of coefficients was found depending on which sample was left out at the time. The distributions found are given in Tables 4 and 5 for the ternary and quaternary systems respectively. These values indicate the sensitivity of the model for data containing relatively large noise amplitudes. The standard deviations for X in the quaternary system are relatively high due to the small and strongly varying concentrations of the constituent trace elements. Of these, the effects of X on the NiK_α and of nickel on the CrK_α appear to have no statistical significance.

TABLE 3

Results for the set of unknowns compared

	Mean error	
	[1]	GA
Chromium	0.0027	0.0016
Nickel	0.0030	0.0025
Iron	0.0045	0.0034

TABLE 4

Distribution of coefficients found for the actual data set (A denotes absorption and B denotes secondary fluorescence effects)

Effect	Type	Coefficient	
		Mean value	S.D.
Cr → FeK_α	A	2.60	0.004
Cr → NiK_α	A	1.31	0.240
Fe → NiK_α	A	2.03	0.011
Fe → CrK_α	B	−0.50	0.094
Ni → FeK_α	B	−0.96	0.078
Ni → CrK_α	B	−0.34	0.025

In the case of the neural network approach, a number of network topologies were tried for all elements. A topology with four neurons in the hidden layer proved to be sufficient for all three elements, and was used in the leave one out method. However, no decrease in performance due to 'overfitting' of the data was detected when larger topologies were used.

The best results for each method of all leave one out experiments are listed for each primary element separately in Table 6. The mean values of the absolute (fractions) and relative (percentages) errors are given. The latter gives rise to very high errors at the low concentrations of chromium and nickel in some of the samples and

TABLE 5

Distribution of coefficients found for the quaternary system (A denotes absorption and B denotes secondary fluorescence effects)

Effect	Type	Coefficient	
		Mean value	S.D.
Cr → FeK_α	A	2.53	0.076
Cr → NiK_α	A	1.42	0.260
Fe → NiK_α	A	1.95	0.095
Fe → CrK_α	B	−0.46	0.020
Ni → FeK_α	B	−1.39	0.181
Ni → CrK_α	B	−0.14	0.125
X → CrK_α	A	1.38	0.216
X → CrK_α	B	1.02	0.248
X → NiK_α	A	0.08	0.281
X → NiK_α	B	0.86	0.277
X → FeK_α	A	0.70	0.131
X → FeK_α	B	0.56	0.256

TABLE 6

Errors of the different methods for chromium, nickel and iron [NN denotes the neural network, SLM the semi-linear method R&H(3) and R&H(4) the ternary and quaternary Rasberry–Heinrich models, and linear the pure linear calibration. Absolute errors are expressed in fractions, relative errors in percentages]

Method	Chromium		Nickel		Iron	
	Abs.	Rel. (%)	Abs.	Rel. (%)	Abs.	Rel.
NN	0.0023	20.63	0.0015	29.03	0.0053	0.68
SLM	0.0022	11.21	0.0020	49.46	0.0064	0.82
R&H (3)	0.0062	11.26	0.0028	32.14	0.0453	5.46
R&H (4)	0.0034	7.89	0.0020	30.98	0.0245	2.99
Linear	0.0087	136.03	0.0039	128.32	0.0363	4.49

serves to illustrate the performance of the various models for these low concentrations. As an illustration of the non-linearity of the measured system, the results of a purely linear calibration are also listed in the tables.

In the case of chromium, the neural net performs slightly worse than the semi-linear model, but much better than the Rasberry–Heinrich model. The latter, however, produces better results for low concentrations as can been seen from the relative errors. In the cases of nickel and iron, the neural net outperforms all other models. Surprising are the relatively bad results of the Rasberry–Heinrich models for iron. This in our opinion can be ascribed to the large levels of noise present in the data set. For all three elements, the application of a quaternary system for the Rasberry–Heinrich model has shown to improve the results a great deal.

Conclusion

In the presented case of determining the composition of steel over large ranges, backpropagation neural networks have shown to be capable of robust multivariate calibration. They can supply accurate quantitative results and have outperformed two other previously known calibration models for this problem. Also, no tendency was detected to overfit the data set when very large topologies were tested or extreme long training times were applied.

The use of the SCG algorithm has led to faster converging networks and has decreased the total amount of calculation time required considerably. Also fast convergence was observed when two hidden layers were used during the search for an optimal topology. In these cases the standard backpropagation algorithm tends to be very slow.

REFERENCES

1 S.D. Rasberry and K.F.J. Heinrich, Anal. Chem., 46 (1974) 81.
2 D.E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning, Addison Wesley, Reading, MA, 1989.
3 J.H. Holland, Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Arbor, MI, 1975.
4 M. Bos, Anal. Chim. Acta, 166 (1984) 261.
5 W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, Numerical Recipes in C: The Art of Scientific Computing, Cambridge University Press, New York, 1989.
6 M. Bos, A. Bos and W.E. van der Linden, Anal. Chim. Acta, 233 (1990) 31.
7 A. Bos, M. Bos and W.E. van der Linden, Anal. Chim. Acta, 256 (1992) 133.
8 D.E. Rumelhart and I.L. McClelland, Parallel Distributed Processing, Vols. 1 and 2, MIT Press, Bradford, 1986.
9 E.M. Johansson, F.U. Dowla and D.M. Goodman, Lawrence Livermore National Laboratory, 1990, Preprint UCRL-JC-104850.
10 M.F. Möller, Neural Networks, in press.