Journal of Thermal Analysis and Calorimetry, Vol. 74 (2003) 811–818

ARTIFICIAL NEURAL NETWORK APPLIED TO SOLID STATE THERMAL DECOMPOSITION

R. C. O. Sebastião, J. P. Braga and M. I. Yoshida*

Departamento de Química, ICEx Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte, MG, Brasil

(Received September 15, 2002; in revised form October 10, 2003)

Abstract

A multi-layer neural network is constructed to describe the thermal decomposition of rhodium acetate. Critical analysis of the residual, trained, interpolated and extrapolated errors, with the number of neurons, indicates the efficiency of the present approach. It was possible, within this framework, to improve the A_n model, with a better correlation between the results. A new value of the activation energy, E_a, and frequency factor, Z, are calculated for the decomposition process. Since the neural network is more precise than a particular model, the calculated values for these quantities are believed to be more precise. The computed values are E_a=194.0 kJ mol^{-1} and Z=5.23·10^{16} s^{-1}. The neural network eliminates the step to decide, among the available models, the one that best fit the data. An agreement up to four significant figures can be achieved even for data not used in the training process, both in the interpolated and extrapolated regions. This method suggests, therefore, an important alternative tool for the experimentalists. The present approach can also be adapted to other systems and to data in two dimensions.

Keywords: multi-layer percepton, neural network, thermal decomposition

Introduction

Theoretical foundation of solid-state decomposition kinetics is based on growth and diffusion of particles, from which several models are proposed to describe the process [1]. A fitting procedure will decide, among the possible models, the one that best adjust to the data. Nevertheless, some models can give errors of comparable magnitude, preventing a correct choice of the model, which is best adapted to the experimental situation. In several occasions, even the minimum residual error is not acceptable for the models available [2]. Some models can also describe better different regions of the decomposition, which makes the selection of the best model only a first approximation to describe the decomposition kinetics.

An alternative approach, using multi-layer neural network [3], will be discussed in the present work. The similarity between the nervous impulse and the time depend-

* Author for Correspondence: jpbraga@oxigenio.qui.ufmg.br

1388–6150/2003/ $ 20.00

Akadémiai Kiadó, Budapest
Kluwer Academic Publishers, Dordrecht

ence of the weight fraction was, in fact, the initial motivation to apply this framework to the thermal decomposition problem. It is, therefore, expected that the neural network will work properly when applied to this process.

This artificial intelligence method also avoids the usage of specific models although the chemical meaning is lost for more than one neuron. Nevertheless, as will be discussed, this approach can be very useful when dealing with experimental data. Before a certain limit on the number of neurons this scheme will converge to desired accuracy, the precision being superior to those given by the usage of any thermal decomposition models previous described in the literature.

The rhodium(II) acetate system [4] will be taken as a model to illustrate the applicability of the neural network approach. The computed values of the mass fraction as a function of time, activation energy and frequency factor are compared with experimental data and previous reported results [4]. Since models of comparable accuracy can give different activation energy values [5], the precision of the above approach will give new, and more reliable, results for this quantity.

Neural network theoretical background

Computer codes that simulate the brain operation are generally termed neural network [3]. These softwares will simulate the transmission of information through synaptic connections and nervous impulse. As in the brain operation, neural networks codes also operate in parallel and, therefore, should be very efficient. It is expected also that neural network have the capability of learning from data, predict information not used in the training process and to establish a relationship between input and output.

For the input data denoted by $i=[i_1, i_2, ...,i_m]^T$ and output $o=[o_1, o_2, ...o_3]^T$ one defines the sum at each neuron

$$s_j = \sum_{k=1}^{1} w_{jk} i_k \tag{1}$$

where w_{jk} simulates the synaptic connection between the neurons k and j.

An activation function, i.e., a function that reproduces the nervous impulse, applied to this sum defines the state of the neuron,

$$y_j = f(s_j) \tag{2}$$

The logarithm sigmoidal activation function, $f(s_j)=1/(1+e^{-s_j})$, was used in the present work. The similarity between this kind of impulse and thermal decomposition model, the A_n model, is obvious at this point.

This process, illustrated for one neuron in Fig. 1, can be used for several neurons and adapted into a multi-layer neural network. The arrangement of the neurons into layers is the neural network architecture.

The learning process is based on minimizing the errors between computed neural network output and the experimental data, which is achieved by adjusting the values of the connections in the hidden and output layers. Since, at the k^{th} layer the neu-

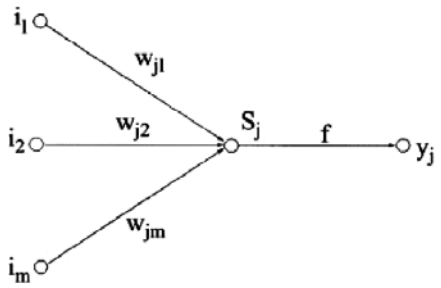

Fig. 1 Structure of a basic neuron

ron state is represented by $y_k(w)$, as a generalization of (2), w being a vector describing the values of the connections, the quadratic error can be written as,

$$E(w)=\sum_k (o-y_k(w))^2=\sum_k (e_k(w))^2 \quad (3)$$

The solution that minimizes the above error can be found by using several optimization techniques, among them, the back-propagation algorithm [6], the recurrent neural network method [7] and the regularized method, to be described here.

Defining the vectors $e(w)=(e_1, e_2, ..., e_m)^T$ and assuming a solution occurs at $w+\delta w$, that is $E(w+\delta w)=0$, the minimum of (3) will be equivalent to solve, in a recurrent way, the linear system [3]

$$J(w)\delta w=-e(w) \quad (4)$$

The quantity J is the Jacobian of the error with respect to the values of the connections and has dimension $m\times n$, m being the number of data and n the number of values of the connections in the neural network architecture. As a consequence of J being ill-conditioned, special techniques to solve (4) have to be applied. The one adopted here was to regularize the Jacobian matrix, implying the new system to be solved will be,

$$(J^TJ+\lambda I)\delta w=-J^Te \quad (5)$$

in which I is the identity matrix and a regularized parameter.

The Levenberg–Marquardt algorithm [8, 9], or the equivalent first order Tikhonov regularization [10, 11], was chosen to find the above solution. The regularized parameter is found by changing its value according to the direction of the total error. The computer calculations described here were performed using the Matlab software [12].

The reference system

The amount of mass lost at time t normalized to the total mass lost, quantity denoted by α, is used to analyze the thermal decomposition process. The neural network efficiency was tested from the data published in [4], at the temperatures 198, 202.5, 203.5, 205.0, 207.0, 209.0 and 210.5°C, where it was found that the A_n model

$$\alpha(t)=\frac{1}{1+e^{-(kt+k_0)}} \tag{6}$$

gives the best fitting to the published data as function of time, t. From the computed values of velocity constant, k, the activation energy was found to be E_a=182.5 kJ mol^{-1} and the frequency factor Z=2.5·10^{15} s^{-1}. These results are to be compared with the ones obtained by the neural network procedure.

Results and discussion

One neuron architecture

The time dependence of α was first investigated for an architecture with one neuron in the hidden layer. Following the rule (1) and (2), and at the hidden layer, the state of the neuron will be $1/[1+e^{-(w_{21}i+w_{20})}]$. At the output, the neuron is activated by a linear function, with the values of the connections w_{32} and w_{30}. The neural network output, within this architecture, is, therefore, represented by,

$$o=\frac{w_{32}}{1+e^{-(w_{21}i+w_{20})}} \tag{7}$$

The residual error, Eq. (3), for the neural network approach and A_n models are given in Fig. 2. The superiority of the neural network approach is already clear at this stage. For one neural architecture the residual error has decreased by a factor of three to seven when compared with the A_n model.

Table 1 Calculated values of the connections for one neuron architecture. Numbers in parenthesis are for power of ten

Temperature/°C	w_{21}/min^{-1}	w_{20}	w_{32}	w_{30}
198.0	9.761 (–4)	–13.95	0.8839	–1.083 (–2)
202.5	1.500 (–3)	–14.12	0.9832	8.060 (–3)
203.5	1.697 (–3)	–14.87	0.9786	5.036 (–3)
205.0	2.267 (–3)	–14.24	0.9652	9.694 (–3)
207.0	2.303 (–3)	–16.94	0.9018	2.026 (–2)
209.0	3.014 (–3)	–21.45	0.9272	2.991 (–2)
210.5	3.524 (–3)	–18.59	0.9186	4.734 (–2)

Activation energy for the decomposition process can, therefore, be recalculated based on the above improved results. The values of the connections for one neuron network and for the various temperatures are given in Table 1. The computed weight of the connections w_{21} and w_{20} represent the constants k and k_0 and are in agreement with the previous published results [4]. In Fig. 3 the logarithm of w_{21} is plotted *vs*. the inverse of absolute temperature, together with the results from [4]. The activation en-

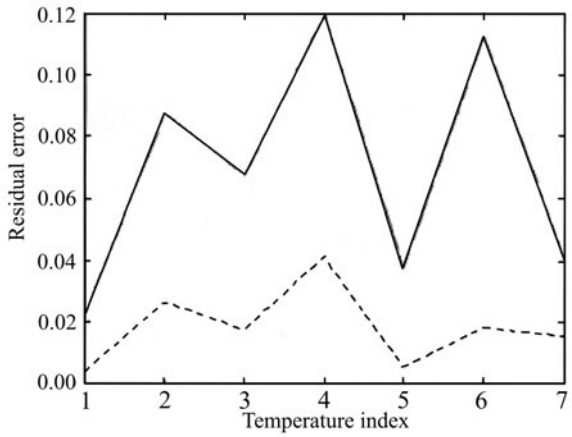

Fig. 2 Errors as a function of temperature. The index for temperatures is 1 for 198°C, 2 for 202.5°C, 3 for 203.5°C, 4 for 205.0°C, 5 for 207.0°C, 6 for 209.0°C, 7 for 210.5°C. Full line are for the A_n model and dashed line for the neural network results

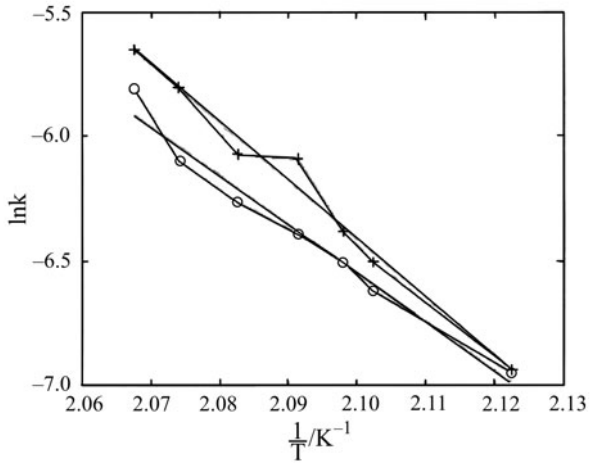

Fig. 3 Logarithm dependence of velocity constant with the inverse of temperature. Crosses are for the the neural network result and circle for the A_n model

ergy and the frequency factor, calculated from the neural network model, are, respectively, E_a=194.0 kJ mol^{-1} and Z=5.23·10^{16} s^{-1}. The results from the neural network, based on the residual error, Fig. 2, are expected to be more reliable.

Equation (7) suggests a reinterpretation for the A_n model. For large decomposition time this model gives α=1. Nevertheless, this does not correspond to the real situation, since this asymptotic values is, in practice, not equal to unity. By introducing the values of the connections w_{32} and w_{30} the neural network procedure corrects this situation.

The correction on time scale can also be interpreted as the introducing of another time scale, known as characteristic time scale, τ, first discussed in [1]. This new time

scale is defined as $t=p\tau$, where p is the probability of the germ nuclei being used up for the new phase. In the present approach this probability corresponds to the value of the connections w_{32}. The time scale in the A_n model is approximately described by the time t.

Together with its physical aspect, the introduction of these two extra variables increase the number of adjustable parameters into the problem, this will improve the correlation between input and output.

Several neurons analysis

As the number of neurons in the hidden layer, N_{neuron}, is increased, the residual error for the trained points, E_{train}, will be reduced. The second column in Table 2 makes clear this situation for T=205°C. Three to four significant figures can be obtained in prediction the decomposition fraction for a given time at the training points. A much better correlation between experimental data and neural network result is achieved in this case. This improvement is evident in Fig. 4 in which the neural network output, for N_{neuron}= 4, and the A_n fitting are confronted with the experimental data.

Table 2 Training interpolated and extrapolated errors for different number of neurons. Numbers in parenthesis are for power of ten

N_{neuron}	E_{train}	E_{inter}	E_{extra}
1	4.121(–2)	5.745(–2)	5.626(–2)
2	4.121(–2)	5.745(–2)	5.626(–2)
4	2.749(–3)	9.339(–3)	8.664(–3)
8	2.959(–3)	5.548(–3)	5.987(–3)
10	3.116(–3)	1.410(–2)	1.498(–2)
11	2.378(–3)	8.685(–2)	2.376(–1)
14	1.412(–1)	1.412(–1)	7.892(–1)

Two other errors are also important to test the robustness of the neural network approach: (a) The interpolated error, E_{inter}, related to points not used in the training process and inside the training region; and (b) the extrapolated error, E_{extra}, for points outside the above region. The third and fourth columns in Table 2 show these two kinds of errors. Although the error for the training points continues to decay, the interpolated and extrapolated errors increase, in an oscillatory way, after N_{neuron}=8. Up to this number of neuron the same precision obtained for the training points can be achieved for interpolated and extrapolated regions.

Further conclusions about the neural network performance can be obtained from Fig. 5. For the same set of training points used in Fig. 4 and also with four neurons in the hidden layer, the neural network framework was able to predict decomposition fractions for time not used in the learning process. Furthermore, the present scheme was also able to predict data, within four significant figures, outside the training region, as shown in the Fig. 5. The neural network powerfulness is more evident at this point.

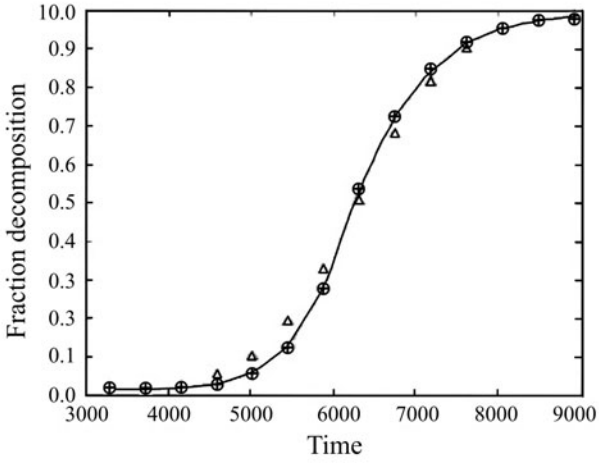

Fig. 4 Time dependence of at T=205°C. Circles are for the neural network result, crosses for the experimental result and triangle for the A_n model

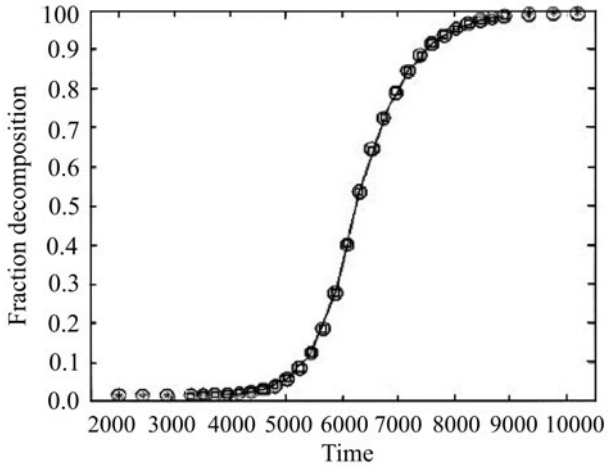

Fig. 5 Time dependence of at T=205°C. Full line are for the training region, squares for the interpolated data, stars for the extrapolated data and circles are for the neural network result. The points used for the training process are as in the previous figure

The superior limit for the number of neurons can be established by finding the region in which the interpolated and extrapolated errors do not decrease as the number of neurons increase, phenomena known as overfitting. For the present case, if more than eight neurons are used, although the results for the trained points will be excellent (four to five significant figures), the ability to predict points will not be reliable.

Conclusions

An alternative approach to describe thermal decomposition data, using neural network, has been discussed in the present work. Architecture with one hidden layer and one to several neurons was used to introduce this alternative scheme. For one neuron the residual error was decrease by a factor of three to seven when compared with the A_n model. As a consequence of this improvement, a better estimation of the activation energy and frequency factor was possible. A prediction of α can be attained within a required accuracy, before the overfitting region has reached. It was possible to predict, with three to four significant figures, data not considered in the training process. This can give valuable informations for the experimentalists.

Non-isothermal data represents no problem for the neural network approach. Within the same framework presented in this paper, a neural network architecture can be constructed to learn about the two dimensional data, $\alpha(t, T)$.

Usage of the neural network suggested also two alterations in the A_n model. The first modification took into account the correct asymptotic value of the mass fraction while the second modification changed the time scale.

Neural networks can be very useful to predict data for a wide range of temperature. For large temperatures, where the decomposition resembles a step function, the neural can still gives excellent results. The general scheme presented here is not restricted to the system under study and can be applied to any other one. This suggests the technique can be used as a powerful routine method to study the chemistry of solid state decompositions.

* * *

The authors would like to thank the agencies CNPq and FAPEMIG for financial support.

References

1 M. Avrami, J. Chem. Phys., 7 (1939) 1103.
2 W. L. Ng, Austr. J. Chem., 28 (1975) 1169.
3 P. De Wilde, Neural Network Models, Springer Verlag, 1995.
4 M. M. Braga, M. I. Yoshida, R. D. Sinisterra and C. F. Carvalho, Thermochim. Acta, 296 (1997) 141.
5 S. Vyazovkin, J. Therm. Anal. Cal., 64 (2001) 829.
6 M. Wesolowski and B. Suchacz, J. Therm. Anal. Cal., 68 (2002) 893.
7 V. C. Viterbo, J. P. Braga, A. P. Braga and M. B. de Almeida, J. Chem. Inf. Comput. Sci., 41 (2001) 309.
8 D. W. Marquardt, J. Soc. Indust Appl. Math., 11 (1963) 431.
9 M. T. Hagan and M. B. Menhaj, IEEE Trans. Neural Network, 5 (1994) 989.
10 A. N. Tikhonov and A. V. Goncharsky, Ill-Posed Problem in the Natural Sciences, Mir, Moscow, 1987.
11 J. P. Braga, J. Math. Chem., 29 (2001)151.
12 D. Hanselman and B. R. Littlefield, Mastering MATLAB 6, Prentice Hall, 2000.