Artificial neural network approach to the evaluation of the coordination geometry in organotin(iv) compounds

Maria Luisa Ganadu,a* Valeria Maida,a Lorenzo Pellerito b and Patrizio Silvi Antonini,c

a Department of Chemistry, University of Sassari, 07100 Sassari, Italy
b Department of Inorganic Chemistry, University of Palermo, Viale delle Scienze 2, Parco d’Orleans, 90128 Palermo, Italy
c Department of Biotechnology, University of Milan - Bicocca, via Emanueli, 12/bis, 20126 Milano, Italy

Received 13th February 1999, Accepted 26th March 1999

Artificial neural networks (ANNs) are a simple and rapid system for pattern recognition. In this study they were used to classify Mössbauer spectra of penta-coordinated and octahedral Sn(iv) complexes. Mössbauer spectra recognition is a lengthy procedure requiring a great deal of experience. The application of a system such as artificial neural networks provides a rapid and accurate method for the correct classification of Mössbauer spectra. As the two categories of spectra are not linearly separable, conventional techniques like principal component analysis (PCA) or perceptron can not be used. A more complex ANN was therefore used to solve this problem. The network was built as a standard three-layer back-propagation network with 256 input neurons, 2 hidden neurons and 1 output neuron and a sigmoidal activation function. The network’s performance was tested with test sets of 10, 20 and 50% of the total number of spectra. The mean square error (MSE) of the different test sets show significant differences. This type of network was able to classify correctly the spectra with an MSE of less than 0.030. Moreover, the network was even able to classify in the appropriate class a spectrum that had been deliberately inverted, demonstrating the ability of ANN to recognize objects affected by noise or distortion.

Introduction

Discriminating between two classes of compounds whose experimental data differ slightly from each other is a challenge which requires great experience and time-consuming procedures.

One of the promising possibilities for solving this type of problem is the application of a method based on artificial neural networks (ANNs) which have been successfully applied for classification and pattern recognition tasks due to their high level of reliability and versatility. Chemical applications of ANNs were recently reviewed by Brown1 and Zupan and Gasteiger.2

Recent applications of the ANNs approach to spectroscopy include classification of aldito3 and identification of carbohydrates on the basis of their 1H-NMR spectra.4,5 An additional attraction of ANNs is their ease of use and the availability of suitable computational packages for building and testing different network architectures. Moreover, using supervised ANNs which ‘learn by example’, it is not necessary to know the exact parameters involved in the interpretation of the particular spectra.6–9

The aim of this study is to present an application of ANNs to the classification of several diorganotin(iv) compounds using the corresponding Mössbauer spectra. Among the diorganotin(iv) derivatives, we chose the following two different classes, namely R2SnCl2 (R = Me, Bu, Ph) and R2SnP2 (where P is a dipeptide or an oligopeptide), which are characterised by two different coordination geometries around the metal centre, and the following peculiar features:

1. The Mössbauer spectra are very similar for compounds belonging to either class, and they are not immediately recognisable;

2. The co-ordination geometry of both classes of organotin(iv) compounds, which has been extensively investigated by X-ray spectroscopy, presents different general trends for the two classes.

In fact, the diorganotin(iv) dihalide (R2SnCl2; R = Me, Bu, Ph) adducts of a series of the electron-donating molecules [such as diphosphonates, diphosphinate and bis(hydroxamate)] have been widely studied because of their biological activity, and, in particular, their potential use as antitumor drugs. X-ray data have shown that in all these complexes the Sn(iv) coordination geometry can be described as a distorted octahedron structure in which the electron-donating ligands chelate the metal ion with a trans configuration.6–9

The diorganotin(iv) and, in particular, dimethyltin(iv), complexes of di-(AlaHis, GlyHis, GlyTyr, HisGly) and oligopeptides (glutathione oxidized, GlyHisGly, GlyTyrHisGly, GluCysGly, ProLeuGlyGly), have been studied by X-ray and Mössbauer spectroscopy to identify the molecular interactions and the structural motifs associated with them. In these compounds, tin(iv) exhibits a distorted trigonal bipyramidal coordination, while the coordinated dipeptides form a network of weak hydrogen bonds that determines the crystal packing of the complexes.10

These two classes of diorganotin(iv) compounds, not related to each other, differ in geometry around a metal centre that is octahedral in the first class, and penta-coordinated in the second one.11,12 Sn Mössbauer spectra can discriminate between the two geometries. The interpretation of Mössbauer spectra requires the use of a number of parameters that can be extracted from the spectra, but great experience and knowledge of Mössbauer spectroscopy is needed to explain the spectrum characteristics. Consequently, since Mössbauer spectra analysis requires great experience and is time-consuming, this can be considered as a typical example of a problem that can be successfully solved with ANNs.

Analyst, 1999, 124, 721–724
Experimental

Sample preparation and instruments

Diorganotin(iv) halides—diphosphonate, -diphosphinate, bis-(hydroxamate) and lipids adducts, and diorganotin peptides complexes, were prepared according to previously reported methods.\textsuperscript{10–13} \textsuperscript{119}Sn Mössbauer spectra were obtained with Ca\textsuperscript{119}SnO\textsubscript{3} (10 mCi, Radiochemical Centre, Amersham, UK) source at room temperature. The absorber samples were powdered and pressed between aluminium foils in a copper sample holder, and maintained at liquid nitrogen temperature in a DN 700 Oxford Instruments (Oxford, UK) cryostat. The 77.3±0.1 K temperature was controlled through an ITC-2 temperature controller from Oxford Instruments. The \textsuperscript{119}Sn concentration was \(\sim 0.5\) mg cm\textsuperscript{-2}.

The source motion was effected with the following Wissenschaftliche Elektronik GmbH apparatus (Starnberg, Germany): a velocity transducer (range \(\pm 10\) m ms\textsuperscript{-1}), an FG-2 function generator and a MR 250 driving unit. Velocity calibration was carried out with an enriched iron foil spectrum (\(57\)Fe = 99.99\%, thickness 0.06 mm, Dupont, Boston, MA, USA) at room temperature using a 57Co source (10 mCi, Dupont) in a palladium matrix, while the zero point of the Döppler velocity was determined, at room temperature, through absorption spectra of natural CaSnO\textsubscript{3} containing 0.5 mg cm\textsuperscript{-2} of \textsuperscript{119}Sn. Finally, a 269 multichannel analyzer (Takes, Ponzanica, Bergamo, Italy) was used and \(5 \times 10^3\) counts were collected for each velocity point.

Artificial neural networks

ANN models\textsuperscript{2,6,12} have aroused a great deal of interest in recent years. They consist of a highly interconnected network of non-linear processing elements, called neurons, since their structure is derived from that of the biological neurons in the brain. We shall concentrate on feed-forward, supervised ANNs. Here the flow of information in the network is one-directional: from the input neurons, which receive a stimulus from the external world, through the hidden neurons, to the output neurons, which in turn provide the response of the network to the input stimulus. The grouping, inside the network, of each of these three distinct types of neurons together, is called a layer. We shall identify an ANN with the notation i–j–k, representing the number of input–hidden–output neurons, respectively, in each corresponding layer. Thus, the set of input stimuli can be viewed as an i-dimensional input vector, which is mapped by the ANN into a k-dimensional output vector. A numerical coefficient (called weight) is associated with each of the connections between any two neurons inside the network. There are non-direct interconnections between input and output neurons. The input to a hidden or output neuron is the sum of the signals coming from the neurons of the previous layer connected to it, weighted by these coefficients. A neuron processes this input via a transfer function. The most interesting results were obtained with nonlinear transfer functions, a typical one being the sigmoid, which approximates the step-function (i.e., the threshold mechanism acting inside the biological neuron) but has good mathematical properties, useful in the learning algorithm.

The way such ANNs work is via learning by example. This means that the coefficients are changed according to an algorithm in such a way that the output from the ANN is as close as possible to the desired output, which has to be known \textit{a priori} by the teacher. In this way, the network learns to associate the input patterns with the desired output. The set of vectors which is shown to the ANN during the training phase is known as the learning set. A target vector is associated with each of these input vectors. The output of the ANN is compared with this target vector, and the mean squared error (MSE) is minimized. MSE is defined as:

\[
\text{MSE} = \frac{1}{P} \sum_{p} \sum_{j} (t_{pj} - o_{pj})^2
\]

where \(t_{pj}\) is the target output neuron \(j\) on pattern \(p\) and \(o_{pj}\) is the actual output computed by neuron \(j\). \(P\) is the total number of patterns presented to the network in the learning process. The learning cycle during which all the training patterns are presented randomly to the network is called an epoch.

The learning algorithm we used to perform this minimization is known as back-propagation. The parameters appearing in this algorithm, as well as the topology of the network (i.e., the number of hidden neurons), were optimized on a trial basis. The learning rate \(\mu\) was set at 0.1. This learning rate is the parameter that determines the amount of weight change at each epoch, so as to optimize the convergence to the minimum MSE. An optimum value for \(\mu\) has to be found in order to improve accuracy, to speed up the convergence process and, at the same time, to prevent the network from being trapped in a local minimum of the multi-dimensional error function.

The weights were set randomly between \(-1.0\) and \(1.0\) before training. The performances of the network were checked on a test set, a subset of all the input vectors different from the learning set. Performances were also monitored during the training phase in order to prevent overtraining of the network, which may occur when the ANN makes an overconstrained fit to the training data, and consequently loses the ability to classify unseen patterns. Overtraining was prevented from stopping the training phase when the MSE was no longer diminishing.

Note that the robustness of ANNs and their ability to generalize and classify incomplete or noisy data, such as those coming from an experimental set-up, were achieved by including this kind of data into the learning set. This is one of the advantages of using ANNs, since there is no need for any algorithmical adjustment to include this experimental variability of the data.

The software used for training and testing the ANN model with our data was SNNS (Stuttgart Neural Network Simulator, University of Stuttgart, Germany). The hardware used was a Silicon Graphics Indigo2 workstation running AIX 3.2.5 and X-Windows.

Spectra and patterns

The first set of data contains the Mössbauer spectra of Sn(iv) with dipptides. In these compounds Sn shows a distorted trigonal bipyramidal coordination (Fig. 1a and Fig. 2a).

The second set of data represents the Mössbauer spectra of Sn(iv) compounds with electron-donating molecules such as diphosphonate, diphosphinate and bis(hydroxamate)s. In this case Sn(iv) geometry can be described as a distorted octahedron (Fig. 1b and 2b).

Due to the symmetry of Mössbauer spectra the 256 value of each folded 512 channel spectrum was chosen to train the network. The 256 values were normalized between 0 and 1. The value 1.0 was added as the output value to the first group of spectra, while the value 0.0 as the output value was added to the second group. In Fig. 1c an example of spectra for each class of compounds is shown.

Three test files were made randomly extracting 10, 20 and 50\% of the complete set of data.

Results and discussion

PCA analysis and perceptron

The aim of this study was the classification of Mössbauer spectra of compounds with two different geometries: octahedral
and penta-coordinated. If the two categories are linearly separable, the problem can be solved using a classical approach like PCA (principal components analysis) or using a simple kind of neural network, known as a perceptron. We therefore tested with PCA in a mathematical procedure which, acting on the variable which characterises the spectrum, extracts a set of components that determine the portion of the spectrum itself in a multi-dimensional variable space, with respect to a hyperplane which separates spectra belonging to the two categories. Thus, the components extracted are linear combinations of the original variables, and altogether they determine the ‘distance’ from the separating plane. The procedure ranks the extracted components in order of importance, and usually the first two or three are enough to classify the spectrum as belonging to one or the other category. Table 1 shows the three principal components for our Mössbauer spectra. They identify a point in a three-dimensional space (since three components were chosen), corresponding to the spectrum. For linearly separable categories, the points corresponding to octahedral spectra should be well separated in space from the points corresponding to penta-coordinated spectra. This was not found in our case, as can be seen from Fig. 2 where the first component is shown vs. the second one, and the first vs. the third one.

These results were confirmed by using a perceptron, i.e., an ANN with only two layers: a 256-neuron input layer and a 1-neuron output layer. Input data were the 256 spectra points as explained above, and the target output was 1 for octahedral spectra and 0 for penta-coordinated spectra.

Both linear and sigmoidal activation functions inside the perceptron were used. Learning rate was set at 0.1, and initial weights were randomly set between $-1$ and $+1$. Table 1 shows the three principal components for penta-coordinated and octahedral complexes of Sn(IV).

<table>
<thead>
<tr>
<th>Penta-coordinated Sn(IV) complexes</th>
<th>Octahedral Sn(IV) complexes</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>Second</td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td>$-1.2034$</td>
<td>$1.1752$</td>
</tr>
<tr>
<td>$-1.2530$</td>
<td>$0.0215$</td>
</tr>
<tr>
<td>$-1.6011$</td>
<td>$-0.9899$</td>
</tr>
<tr>
<td>$-1.5136$</td>
<td>$-0.4955$</td>
</tr>
<tr>
<td>$0.1416$</td>
<td>$0.1810$</td>
</tr>
<tr>
<td>$0.2135$</td>
<td>$-0.0526$</td>
</tr>
<tr>
<td>$3.2708$</td>
<td>$-2.9326$</td>
</tr>
<tr>
<td>$4.2134$</td>
<td>$0.7594$</td>
</tr>
<tr>
<td>$3.1266$</td>
<td>$-1.4321$</td>
</tr>
<tr>
<td>$0.6323$</td>
<td>$-1.4843$</td>
</tr>
<tr>
<td>$-0.6973$</td>
<td>$-0.1581$</td>
</tr>
<tr>
<td>$-0.4285$</td>
<td>$0.2687$</td>
</tr>
<tr>
<td>$1.7979$</td>
<td>$1.2330$</td>
</tr>
</tbody>
</table>

Fig. 1 Normalized Mössbauer spectra of complexes of Sn(IV), a, in penta-coordinated geometry (P), b, octahedral geometry (O), and c, spectrum of each class of Sn(IV) compounds.

Fig. 2 Results of PCA analysis. a, Plots of the first principal component vs. the second one for both penta-coordinated (•) and octahedral (○) complexes. b, Plots of the first principal component vs. the third one for the same complexes as in a.

Fig. 3 Trend of the mean square error on the training set with the perceptron and with a back propagation ANN. Curves a, b and c refer to ANNs with 10, 20 and 50% test files, respectively. MSE is 0 for the 10 and 20% test files, and 0.03 for the 50% one.
MSE, constructed on the test set, as a function of the learning epochs. Both 256-1 perceptrons, with linear and sigmoidal activation functions, were unable to discriminate properly between the two categories, as demonstrated by the fact that the MSE stayed high no matter how many training epochs we used.\textsuperscript{15}

The failure of these approaches suggested that neither could the problem be solved by this method, and nor was it linearly separable. Therefore, a more sophisticated architecture for ANNs must be used.

**Three layer network**

A three layer feed-forward network, with full connection topology between layers, sigmoid type non-linear activation function and an error back-propagation learning rule, were then used. The topology is a standard three-layer feed-forward network with 256 input neurons, normalised between 0 and 1, two hidden neurons and one output neuron.

In our application, the input neurons receive the spectra as a vector whose elements represent half of the elements of the original spectrum due to its symmetry. Input and output data were defined as for the perceptron. Initial weights randomly varied between $+0.1$ and $-0.1$ and the learning rate was set at 0.1. The performance of the last approach was checked on a test set (a subset of all the available data) chosen at random by extracting from 10 to 50% of total files.

As a result of the network operation, the output neuron had a value greater than 0.9 for all the patterns presented to the ANN belonging to the first group, while it was consistently below 0.1 for all the patterns belonging to the second group. This indicates that the 256-2-1 ANN was efficient for the test sample and thus capable of discriminating the two categories.

Fig. 3 shows the MSE, computed on the test set, as a function of the learning epochs.

In order to test the capabilities of the network, we tried several test sets, made up by 10, 20 and 50% of all the available spectra. The size of the test set influences the performance of the network, since a larger test set implies a smaller learning set, and consequently poorer quality in the learning data set, as well as a slower approach to the minimum. However, this does not impair the ability of the network to generalize (\textit{i.e.}, to classify unseen patterns) but is reflected in greater uncertainty in the classification.

As mentioned in the introduction, a great advantage of ANNs over the conventional algorithm is the ability to classify objects even if affected by noise as distortion. This was verified in our case by feeding the network with a deliberately inverted spectrum (but keeping other characteristics unchanged). Even in this extreme case, the spectrum was correctly classified.

**Conclusions**

In this study we presented a simple and quick method to evaluate the geometry of Sn($\text{iv}$) compounds through the classification of Mössbauer spectra: the use of artificial neural networks. Interpreting Mössbauer spectra is a delicate, time-consuming task. Linear separability in the classification of these spectra does not apply, thus conventional techniques like PCA or perceptron cannot be used.

However, a more complex ANN could solve the problem satisfactorily. This approach can be easily adapted to the classification of other kinds of Mössbauer spectra.

**Acknowledgement**

The study was supported by Regione Autonoma Sardegna (RAS).

**References**