Journal of Chromatography, 112 (1975) 297–300

CHROM. 8446

ANALYSIS OF PYROLYSIS GAS CHROMATOGRAMS USING THE LINEAR LEARNING MACHINE METHOD

E. KÜLLIK, M. KALJURAND and M. KOEL
Institute of Chemistry, Academy of Sciences of the Estonian SSR, Tallinn (U.S.S.R.)

SUMMARY

The linear learning machine method for the analysis of pyrolysis gas chromatograms of polymers has been investigated. Classifiers have been evaluated in order to identify different types of molecular groups in polymer fibres with predictive abilities between 70 and 90%. Some aspects of the use of this method for processing pyrograms are discussed.

INTRODUCTION

Pyrolysis gas chromatography is a powerful analytical technique for the analysis of polymers. Chromatograms obtained by the pyrolysis technique are used mainly for the identification of substances. This method is known as the fingerprint method and sometimes it is compared with the mass spectrometric method. Its disadvantage is that the real chromatogram of the same sample must be obtained in order to be able to identify unknown substances. Also, the outer form of pyrograms depends to a great extent on the pyrolysis and conditions of chromatographic analysis. Such a situation suggests the need for the compilation of a cataloque of pyrograms and the application of standard conditions for any class of substances analyzed by gas chromatography. Several attempts have been made, but so far there has been no progress and such catalogues exist only in laboratories where specific problems are being studied. Success in obtaining identical results in other laboratories also depends on the possibility of repeating the same conditions for the pyrolysis and gas chromatographic analysis of thermal degradation products.

On the other hand, the different pyrolysis gas chromatograms resulting from substances with different structures are useful for identification purposes, but the problems involved in the determination of the structures of substances by using gas chromatograms has not yet been solved.

Such a situation compelled us to find new possibilities for determining the structure of an unknown polymer according to its pyrogram. Recently, a spread pattern recognition method has been shown to offer many possibilities for solving these problems[1–4]. The aim of this paper was to study the application of the pattern recognition method in pyrolysis gas chromatographic analysis.

THEORETICAL

The following treatment is based on geometrical concepts and some of the terminology used is derived from Nilsson's book[5]. Let a certain set of measurements $x_1, x_2, \ldots, x_n$ characterize a substance or a polymer. The numbers may represent mass, infrared (IR) or nuclear magnetic resonance (NMR) spectra, pyrolysis gas chromatograms, etc., in digital form.

These sets of measurements may be considered as a pattern vector, $\vec{X}$, in n-dimensional pattern space that extends from the origin to the point $x_1, x_2, \ldots, x_n$. An $(n+1)$ component, the value of which is always 1, is added to the original pattern vector $\vec{X}$:

$$\vec{X} = (x_1, x_2, \ldots, x_n, 1) \tag{1}$$

Let the substance or polymer belong to one of two classes such that one of the classes may have a certain property, while the second class has no such property. Let another set of parameters, $w_1, w_2, \ldots, w_{n+1}$, be derived in some way. Then, we find the value S given by

$$S = \sum_{i=1}^{n+1} w_i x_i \tag{2}$$

The substance belongs to the first class if $S > 0$, otherwise it belongs to the second class. This set of numbers may be considered as a vector $\vec{W}$ in an $(n+1)$-dimensional space and is the so-called weight vector, given by

$$\vec{W} = (w_1, w_2, \ldots, w_{n+1}) \tag{3}$$

Thus it can be seen that the existence of a property in a substance can be determined according to eqn. 2 if the corresponding weight vector $\vec{W}$ is known. In order to find $\vec{W}$, the algorithm known as the teaching with the error correction can be used. The algorithm consists in starting from the arbitrary weight vector $\vec{W}$ and from the known set of vectors $\vec{X}_1, \vec{X}_2, \ldots, \vec{X}_l, \ldots, \vec{X}_l$ (*i.e.*, for every $\vec{X}_l$ vector we know the class to which it belongs), where l represents the training set size. If we obtain the wrong classification, we can correct the vector $\vec{W}$ according to eqn. 4:

$$\vec{W}' = \vec{W} + c\vec{X}_l \tag{4}$$

where $\vec{W}'$ is a corrected weight vector and c is the increment coefficient. The process is repeated until the weight vector classifies all the members of the training set correctly.

We should point out that for the successful application of the algorithm observed, it is necessary for patterns of different classes to fill different regions in an n-dimensional space, which cannot overlap. This condition is not known in advance and its validity appears during the application of the algorithm.

TABLE I

NUMBERS OF FIBRES IN TRAINING AND PREDICTION SETS

Fibres	*Number of fibres in training set*	*Number of fibres in prediction set*
Polyamide	5	9
Polyester	5	4
Polyurethane	5	2
Polyvinyl	7	6
Polyacrylonitrile	8	12
Total	30	33

INITIAL DATA

In this instance, the set of measurements was obtained on special coded pyrolysis gas chromatograms. Pyrograms of textile fibres given in the literature were used as initial data[6]. The set of 63 pyrograms consisted mainly of gas chromatograms obtained by the pyrolysis of polyamide, polyester, polyurethane, polyvinyl and polyacrylonitrile fibres. Thirty pyrograms comprised the training set and the remaining 33 pyrograms were used in order to determine the prediction ability of the weight vector. All of these pyrograms form the so-called prediction set and are given in Table I. The prediction ability refers to the ability of the pattern to classify correctly the patterns that were not the members of the training set.

PRESENTATION OF PYROGRAMS TO THE COMPUTER

The pyrolysis gas chromatograms used in this work were obtained under similar conditions from recorder traces. Each pyrogram was divided into 39 zones (*i.e.*, $n = 39$). The most intensive peak was chosen from every zone and was repre-

TABLE II

RESULTS OF PREDICTION AND RECOGNITION

Class	*Dimension*						
	10			*20*		*40*	
	No. of feedbacks	*% of recognition*	*% of prediction*	*No. of feedbacks*	*% of prediction*	*No. of feedbacks*	*% of prediction*
$-CH_2-CHR-$	200	86.6	72.7	12	81.8	8	81.8
$-(CH_2)_n-$ ($n = 2$–4)	200	79.9	60.6	22	84.8	8	75.7
$-(CH_2)_n-$ ($n \geq 5$)	200	63.3	66.6	20	93.9	9	84.8
$-O-$	200	63.3	45.4	40	69.6	11	78.7
$-CO-$	200	86.6	72.7	12	72.7	8	81.8
$-Cl$	200	83.3	84.8	16	87.8	12	87.8
$-N=$	200	73.3	66.6	36	75.7	10	75.7
$-C\equiv N$	48	100.0	96.9	9	84.8	2	90.9
$-CO-NH-$	200	66.6	66.6	20	84.8	9	84.8
$-C_6H_4-$	200	73.3	69.6	11	90.9	8	84.8
$-CO-O-$	200	79.9	60.6	22	84.8	8	75.7

sented by a digit in the range 0–9. The width of the zone increased linearly with increasing zone number, which was necessary for considering the dependence of the widening of chromatographic peaks on retention time. Dimension reduction is a problem of interest in the pattern recognition method. If a certain feature of one pattern is the same as that of the other, then this feature cannot be used for differentiating the objects. Leaving out features with only small changes, two new pattern vector sets with dimensions of 10 and 20 were formed. The coded pyrograms were punched on to paper tape and calculations were carried out on a Videoton 1010B computer using programs written in Fortran.

RESULTS AND DISCUSSION

The results are shown in Table II. It can be seen that the 20- and 40-dimensional training sets form linearly separated sub-sets. The prediction ability of one set is equal to that of the other and is high. At the same time, the 10-dimensional iteration process did not converge (except for the $-C\equiv N$ class) even after the 200th iteration cycle. For the 10-dimensional vectors, the recognition was calculated. The recognition shows the part of vectors correctly classified in the training set. Both recognition and prediction ability are low in the 10-dimension case.

In these experiments, the best results were obtained by using the 20-dimensional weight vector.

The high prediction ability and the formation of linearly separated sets show that this method can be used successfully for the analysis of pyrolysis gas chromatograms. Although we cannot say that a substance corresponds to a particular pyrogram, as when using file searching, we can determine the structural elements of an unknown substance according to eqn. 2 using all the weight vectors and evaluate the structure of a substance according to its pyrogram. The essential difference between the two methods lies in the fact that a file of pyrograms must have tens of thousands of members, while the learning machine method can be used with hundreds of members of a training set. The use of larger training sets, which contain 100–300 pyrograms, makes it possible to generalize the present results. The computer is no longer needed when making decisions according to eqn. 2 and a desk calculator can be used. The information obtained by the linear learning machine method can be of help in determining the structure of an unknown polymer.

REFERENCES

1 T. L. Isenhour and P. C. Jurs, *Anal. Chem.*, 43 (1971) 20A.
2 I. A. Vink, W. B. Heerma, I. P. Kamerling and J. F. G. Viliegenthart, *Org. Mass. Spektrom.*, 9 (1974) 536.
3 P. D. Tunnicliff and P. A. Wadsworth, *Anal. Chem.*, 45 (1973) 12.
4 K. Warmuza, H. Rotter and P. Krenmayr, *Chromatographia*, 7 (1974) 522.
5 N. J. Nilsson, *Learning Machines*, McGraw-Hill, New York, 1965.
6 O. Kirret and E. Küllik, *Lenziger Ber.*, 36 (1974) 222.