Chart I

cis-Butadiene		*trans*-Butadiene	
$\|k\rangle$ = $^1B_1$	−154.390 au	$\|k\rangle$ = ^1A_u	−154.395 au
$^1B_2$	−154.421 au	^1B_u	−154.397 au
$\|n\rangle$ = $^1A_1$	−154.489 au	$\|n\rangle$ = ^1A_g	−154.569 au
$\|o\rangle$ = $^1A_1$	−154.766 au	$\|o\rangle$ = ^1A_g	−154.722 au

In order to relate the selection rules to the conrotatory and disrotatory Woodward–Hoffmann displacements we focus our attention on the possible twist of the CH_2 terminal groups.

In the case of *cis*-butadiene the symmetries of the vibrations associated with the $^1B_1$ and $^1B_2$ excited states are respectively B_1 and B_2. The B_1 vibration does not provide any twist of the CH_2 groups whereas the B_2 one displays a disrotatory behavior (Figure 4a). This photochemical disrotatory motion is in accord with the Woodward–Hoffmann rules and produces cyclobutene.

In the *trans*-butadiene case the vibrational symmetries associated with the ^1B_u and ^1A_u excited states are respectively B_u and A_u. Here again B_u does not involve any twisting motion whereas the A_u vibration is a disrotatory one (including also the in-phase out-of-plane motion of the central hydrogen atoms (Figure 4b). These two motions distort the highest occupied MO in such a way that the increased overlaps (arrows in Figure 5) lead to the formation of bicyclo[1.1.0]butane (**2**).

This reaction may be a slow concerted one or a two-step reaction involving the intermediate **1**.[26]

Acknowledgment. This work was supported by Research Grant No. GM 13468 from the National Institutes of Health.

(26) R. Srinivasan, *J. Amer. Chem. Soc.*, **90**, 4498 (1968).

Pattern Recognition.[1] A Powerful Approach to Interpreting Chemical Data

B. R. Kowalski and C. F. Bender*

Contribution from the General Chemistry Division, Lawrence Livermore Laboratory, University of California, Livermore, California 94550. *Received December 18, 1971*

Abstract: Pattern recognition is a newly developing branch of artificial intelligence that shows a great deal of promise in providing a generalized approach to solutions of a large class of data analysis problems in experimental chemistry. A general statement of the problem is: can an obscure property of a collection of objects (elements, compounds, mixtures, etc.) be detected and/or predicted using indirect measurements made on the objects? One particular method within the realm of pattern recognition, the learning machine, has been successfully applied to spectroscopic data for direct detection of molecular structural units. This paper introduces pattern recognition in a much broader scope. Using a synthetic data base and a data base of chemical interest, the major approaches within pattern recognition are examined. One method representing each approach is applied to the two fundamentally different data sets, first to compare the results, but also to illustrate the far-reaching problem solving capability.

A large amount of experimental science deals with predicting properties of objects which are not directly measurable. In chemistry, the objects range from pure elements or compounds to complicated industrial and natural products. The properties can be fundamental, such as atomic or molecular structure, or less fundamental, such as reactivity, permeability, absorptivity, etc. All too often, these properties are not directly measurable and must be found using experimental measurements which are known to be related, in some way, to the sought-for property. In some cases a theoretical relationship between measurements and the property is used. A few simple but common examples serve to clarify this point. Emission spectrometry does not provide a direct measure of atomic composition (few methods do) but rather a measure of the wavelengths of light emitted when a sample is "pumped" with energy. The mathematics of atomic theory provide the connection between combinations of various wavelengths and the structure of atoms. Along the same lines, nmr spectrometry does not provide a direct measure of molecular structure but rather a measure of how isotopes are perturbed under various experimental conditions. Group theory provides the connection between nmr parameters and molecular structure.

To proceed, let us consider a less fundamental property of chemical compounds, reactivity. We will assume that one is faced with the problem of predicting the reactivity within a very large number of samples (compounds or mixtures of compounds). There are three methods of determining whether or not two compounds will react in a prescribed manner. The first and most obvious is the direct determination method consisting of adding one to the other under the desired conditions of temperature and pressure. The next method, herein called the theoretical method, is to study bonding possibilities of the molecules taking into consideration such things as orbital symmetry, steric hindrance, etc. Although these two methods are the most desirable, they may not be feasible. Direct methods may be prohibitively expensive, time con-

(1) Work performed under the auspices of the U. S. Atomic Energy Commission.

suming, or dangerous. A theoretical approach is best in these cases but may be impossible since the proposed reactants might be complicated mixtures of unknown composition (*i.e.*, natural products). The scientist is usually forced into using a third method: the educated guess. This approach should not be thought of as being unscientific. The human (especially the scientist) is very efficient at learning from experience and is capable of using a high-order logic in drawing conclusions. Very little research has been done to systematize this third process in such a way as to provide general solutions to problems. A general statement of the problem is: given a set of objects and a list of measurements made on these objects, is it possible to find and/or predict a property of the objects that is not directly measurable but is known to be related to the measurement *via* some *unknown* relationship?

A new field has recently emerged from applied mathematics which shows a great deal of promise in solving this class of problems. It is called pattern recognition[2-4] and employs some very unique techniques of problem solving. Pattern recognition techniques were originally used to solve data processing problems in a number of diverse areas. These areas include handwritten and printed alphanumeric character recognition, weather prediction, medical diagnosis, speech analysis, and many others. Recently, a number of papers have entered the chemical literature describing applications of one particular pattern recognition method, the linear learning machine, to the analysis of various types of spectroscopic data.[5-10] The titles of these papers have included such names as "Pattern Recognition" and "Machine Intelligence" but the papers are really applications of nonparametric learning machines employing feedback learning and a threshold log unit.[11]

This paper introduces the field of pattern recognition to the chemical literature in a much broader scope. It can be used as an introduction to the field in that it touches on the main branches of this new discipline by using one method from each branch to analyze chemical data. No attempt will be made to introduce the reader to the broad field of artificial intelligence. Suffice to say that pattern recognition is a subset of artificial intelligence. Excellent material is available to those who wish to learn about the other branches of artificial intelligence.[12-14]

(2) (a) J. M. Mendel and K. S. Fu, Ed., "Adaptive Learning and Pattern Recognition Systems," Academic Press, New York, N. Y., 1970; (b) K. S. Fu, "Sequential Methods in Pattern Recognition and Machine Learning," Academic Press, New York, N. Y., 1968.

(3) S. Watanabe, Ed., "Methodologies of Pattern Recognition," Academic Press, New York, N. Y., 1969.

(4) G. S. Sebestyen, "Decision-Making Processes in Pattern Recognition," MacMillian, New York, N. Y., 1962.

(5) P. C. Jurs, B. R. Kowalski, and T. L. Isenhour, *Anal. Chem.*, **41**, 21 (1969).

(6) B. R. Kowalski, P. C. Jurs, T. L. Isenhour, and C. N. Reilley, *ibid.*, **41**, 1945 (1969).

(7) B. R. Kowalski and C. A. Reilly, *J. Phys. Chem.*, **75**, 1402 (1971).

(8) L. E. Wangen and T. L. Isenhour, *Anal. Chem.*, **42**, 737 (1970).

(9) L. B. Sybrandt and S. P. Perone, *ibid.*, **43**, 382 (1971).

(10) T. L. Isenhour and P. C. Jurs, *ibid.*, **43**, 20A (1971); also see references within.

(11) N. J. Nilsson, "Learning Machines," McGraw-Hill, New York, N. Y., 1965.

(12) E. A. Feigenbaum and J. Feldman, Ed., "Computers and Thought," McGraw-Hill, New York, N. Y., 1963.

(13) B. Meltzer and D. Michie, Ed., "Machine Intelligence 4," American Elsevier, New York, N. Y., 1969.

(14) M. Minsky, *Proc. IRE*, **49**, 8 (1961).

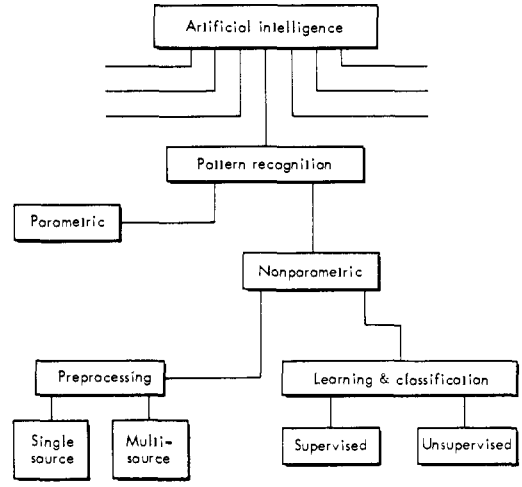

Figure 1. Functional breakdown of pattern recognition techniques.

Pattern Recognition Approach

Figure 1 shows a functional breakdown of the field of pattern recognition. For most chemical applications, the underlying statistics of the properties or classes to be found are usually not known. Parametric methods of pattern recognition assume that probability density functions are known or can be estimated. Bayes strategies[15] are employed in the learning and decision process. While this branch of pattern recognition might possibly be very fruitful in the future, the problems associated with practical applications are many. Therefore, this paper will be concerned with the nonparametric branch of pattern recognition which makes no assumption about the underlying statistical distribution of the data.

As stated earlier, the goal is to recognize an obscure property in a collection of objects from indirect measurements made on the objects. Two questions must be answered at the outset. Unfortunately they are interrelated. (1) What must be learned from the objects? (2) Are the data (measurements) in the correct form? These questions lead to the first branching under nonparametric pattern recognition (Figure 1).

Preprocessing is numerically operating on the data (measurements) in order to change the representation of the information contained in the data. Learning and classification proceed once the transformation is completed and lead directly to the desired results. To illustrate the use of the measurements a geometric description of pattern recognition will be used.

Consider the objects as points in an n-dimensional space, where n is equal to the number of measurements made on each object. Of course, the same measurements must be made on each object. The measurements are the coordinates of each point in n space. The distance between two points in the n space, an excellent measure of their "likeness," is the simple Euclidean distance. (There are other types of distances depending on the chosen metric, and hence other types of similarity measures, but they will not be discussed here.) It is assumed that nearness in space between two points is a good measure of similarity between the corresponding objects and that a sufficient estimate of the

(15) T. W. Anderson, "An Introduction to Multivariate Statistical Analysis," Wiley, New York, N. Y., 1958, p 130.

similarity between object X_i and X_j is

$$d_{ij} = \left(\sum_{k=1}^{M} (X_{ik} - X_{jk})^2 \right)^{1/2} \quad (1)$$

where the summation is over the measurements.

Actually, d_{ij} is a reciprocal similarity measure because objects are more alike as d_{ij} goes to zero. To remedy this, a similarity measure is defined as

$$S_{ij} = 1 - d_{ij}/\mathrm{MAX}(d_{ij}) \quad (2)$$

where $\mathrm{MAX}(d_{ij})$ is the largest interpoint distance. For this function, the most unlike objects give $S_{ij} = 0$ and identical objects give $S_{ij} = 1$. Similarity measures are an extremely important part of pattern recognition.

Classification and learning methods operate directly on the n space in either of two modes. The first has been termed supervised learning. Supervised learning means that some of the points in the n space are "tagged" with a *known* classification (these points comprise a *training set*) and the primary objectives are to develop a rule which classifies these points correctly and then apply the same rule for classification of unknown points. The second mode is called unsupervised learning. Here, the objective is to find realistic densities or clusters of points in n space which reflect the possible existence of meaningful interrelationships. Thus in this mode, there is no training set.

Both of these modes, but especially the latter, are dependent on the absolute magnitudes of the measurements. If, for example, in a two-dimensional case, one of the measurements is bond length (in Å) and the other is boiling point (in °C), the points may have a small variance in the axis of the former measurement and a large variance in the latter. It is obvious that scaling is necessary. Scaling is but one type of preprocessing that might be necessary to produce realistic results in the classification and learning stage of a pattern recognition study.

Since visual examination of points in n space (when greater than three dimensional) is not possible, computers are used to analyze the data. Fortunately, the concepts of Euclidean geometry (distance, angles, etc.) hold true in higher dimensional spaces. Besides implementing the preprocessing methods mentioned above, the computer can be used to (1) generate an approximation of the points (n space) in a two-dimensional space so that visual examination is possible, (2) find clusters or densities of points, and (3) classify unknown points according to their nearness (in some sense) to known points. These computer tasks will be discussed in detail individually in later sections of this paper. Discussions of methodology will be aided by two example data sets. One set consists of artifically generated points and most clearly shows how the methods work. The second data set is of chemical interest.

Data

Artificial data, herein referred to as set 1, were generated by computer and contained 75 points in three-dimensional space. The data can be thought of as coming from 75 objects where three measurements (X, Y, and Z) were made on each object. Actually, set 1 was drawn from three randomly generated Gaussian distributions. The three distributions were displaced by adding constants to the coordinates in order to make them well separated. Twenty of the points from each of the three distributions have been labeled as "knowns" leaving a total of 15 unknowns. The problem in the first case is twofold: first to detect the presence of these three groups and, second, to classify unknown points into one of these three groups. It should be remembered that a three-dimensional problem was chosen purely for demonstration. The methods to be discussed are capable of handling more difficult problems.

There was a great deal of difficulty deciding on the second data set, set 2. Requirements were that the problem should use data of general interest to chemists, not have a trivial solution, but at the same time not be overly complicated. The most important requirement was that the problem, and the data used in solving the problem, should be used only as an example and not in any way shadow the true goal of this paper which is to introduce the field of pattern recognition to chemical data processing. It would be difficult to find anything more basic to chemistry than the chemical elements. It was therefore decided to use the elements as the objects for the study. The problem is hypothetical in nature but uses real data that were readily available.

The "Periodic Table of the Elements" published by Sargent-Welch Co.[16] contains a wealth of chemical information. Among the properties listed in this periodic table is whether the representative oxide (higher valence) of each element is acidic, amphoteric, or basic. For the pattern recognition problem, six properties of the elements were used: (1) most important valence, (2) melting point, (3) covalent radius, (4) ionic radius, (5) electronegativity, and (6) ΔH of fusion. All of the properties could not be found for all of the elements so some elements were eliminated. Also, the inert gases were not used for obvious reasons and hydrogen was eliminated because its oxide (H_2O) was presumably the solvent system used to determine the given acidity and basicity. In all there were 68 elements with 27 in the "basic" class, 21 in the "acidic" class, and 20 amphoterics. The objectives of the problem were to see whether the acidic class could be separated from the basic class using the six properties and, if so, to decide how to classify the amphoterics. It should be mentioned that *none of the properties could effect the separations when used alone.* Table I lists the elements used and their identification numbers.

It is important that the reader see beyond these applications and discover the relevance of what is presented here to the wide variety of possible applications in chemical research.

Preprocessing

As was mentioned earlier, preprocessing involves actually changing the structure of points in the n space. Even though there are several preprocessing methods to choose from,[17-19] this branch of pattern recognition has not received much attention until recently and

(16) "Periodic Table of the Elements," Sargent-Welch Scientific Co., Skokie, Ill., 1968.

(17) K. S. Fu, P. J. Min, and T. J. Li, *IEEE Trans. Syst. Sci. Cybernetics*, **SSC-6**, 33 (1970).

(18) M. D. Levine, *Proc. IEEE*, **57**, 1391 (1969).

(19) Institute of Electrical and Electronics Engineers Conference Record of the Symposium on Feature Extraction and Selection in Pattern Recognition, Argonne National Laboratory, Oct 1970.

Table I. Identification Numbers of the Elements Used

1.	Lithium	24.	Antimony	47.	Bismuth
2.	Boron	25.	Tellurium	48.	Thorium
3.	Nitrogen	26.	Iodine	49.	Beryllium
4.	Sodium	27.	Cesium	50.	Aluminum
5.	Magnesium	28.	Barium	51.	Silicon
6.	Phosphorus	29.	Lanthanum	52.	Titanium
7.	Sulfur	30.	Cerium	53.	Vanadium
8.	Chlorine	31.	Praseodymium	54.	Iron
9.	Potassium	32.	Neodymium	55.	Cobalt
10.	Calcium	33.	Samarium	56.	Zinc
11.	Scandium	34.	Europium	57.	Gallium
12.	Chromium	35.	Gadolinium	58.	Germanium
13.	Manganese	36.	Terbium	59.	Zirconium
14.	Arsenic	37.	Dysprosium	60.	Rhodium
15.	Selenium	38.	Erbium	61.	Silver
16.	Bromine	39.	Thulium	62.	Indium
17.	Rubidium	40.	Lutetium	63.	Tin
18.	Strontium	41.	Tantalum	64.	Hafnium
19.	Yttrium	42.	Tungsten	65.	Gold
20.	Niobium	43.	Rhenium	66.	Lead
21.	Molybdenum	44.	Osmium	67.	Polonium
22.	Ruthenium	45.	Mercury	68.	Uranium
23.	Cadmium	46.	Thallium		

therefore a general theoretical approach is nonexistent. It is hoped that this situation will soon be remedied.

It is of the utmost importance that the ratio (R) of the number of objects to the number of measurements used for learning and classification be as large as possible. Sammon, *et al.*,[20] have found for the two class problem that the error rate is a monotonically decreasing function of this ratio. They found, for $R \leq 2$, that geometrically perfect, but possibly meaningless, results could almost always be obtained on design training data sets. Ratios of $R > 3$ were found to be tolerable and ratios of $R > 10$ were most desirable. In the present work, $R = 20$ for set 1 and $R = 8$ for set 2.

When an unsupervised approach is used, almost no preprocessing is justified. If the classifications are not known *a priori*, then very little can be done to improve classification performance. About the only action that can be justified is scaling. When measurements of different units are compared, a weighting of the measurements with the largest absolute values is inadvertently applied. The two-dimensional example cited above (bond length and boiling point) is just such a case. To scale the measurements so that they each have an equal weight and therefore an equal effect on the application, autoscaling is applied. The measurements are scaled so that they each have a mean of zero and unit variance. The kth coordinate of the ith point then becomes

$$Y'_{ik} = (Y_{ik} - \overline{Y_k})/\sigma_k \tag{3}$$

where

$$\bar{Y}_k = 1/N \sum_{i=1}^{N} Y_{ik} \tag{4}$$

and

$$\sigma_k^2 = \sum_{i=1}^{N} (Y_{ik} - \bar{Y}_k)^2 \tag{5}$$

The second kind of preprocessing that was used in this study is simply a weighting of the variables; this can only be used for supervised learning. Since the classifications are known, each coordinate can be weighted according to its relative importance in effecting a separation of the known classes. One example of this technique, classification weighting, is the ratio of the interclass variance to the intraclass variance. It is applied to the coordinates *after* they have been autoscaled.

The new value is

$$Y'_{ik} = \phi_k Y_{ik} \tag{6}$$

where

$$\phi_k = \frac{\sum_a \sum_b P_a P_b X_{ab}^k}{\sum_c P_c X_{cc}^k} \tag{7}$$

and

$$X_{ab}^k = \sum_{i,j} (Y_{ik}^a - Y_{jk}^b)^2 \tag{8}$$

P_a is the simple probability of a point being in class a and is estimated from the data set (number in class/number in data set). X_{ab}^k is the interclass variance and the summation is over all two-point combinations not in the same class. X_{cc}^k is the intraclass variance and the summation is over all two point combinations where both points are in the same class. ϕ_k is calculated for each of the measurements. Larger values of ϕ_k indicate the relative importance of the kth variable. Set 2 was autoscaled and weighted, but set 1 underwent no preprocessing.

It is convenient to refer to the coordinates of the new n space generated by preprocessing as features. These features can be, and often are, quite different from the original measurements. They may be linear or nonlinear (exponential, etc) combinations of some or all of the measurements, depending on the particular preprocessing method used.

The above discussion pertains to what may be called multisource data (Figure 1) meaning that the measurements are made using a variety of instruments. Single-source data (Figure 1) are here defined as coming from the same instrument. Spectroscopy is a good example of the latter because by digitizing a spectrum, several measurements are obtained from one spectrum. The early application of learning machines to mass,[5] infrared,[6] nmr,[7] and γ-ray spectroscopy[8] serves as examples of the application of one type of pattern recognition to single-source data. Preprocessing in single-source studies such as these are many and have included transformations such as autocorrelation[7] and Fourier.[21] These transformations may be necessary to achieve the goals of the particular study. This subject will be treated in greater detail in a subsequent paper.

Visual Display by Mapping

It would be erroneous to infer that pattern recognition removes the scientist from data analysis. Man–machine interaction is currently in vogue in chemistry and other fields for a good reason: man is the best pattern recognizer known today. The difficulty comes when the measurements and/or the objects are many.

(20) Proceedings of the 1970 Institute of Electrical and Electronic Engineers, Symposium on Adaptive Processes, p IX.2.1, University of Texas at Austin, Dec 1970.

(21) L. E. Wangen, N. M. Frew, T. L. Isenhour, and P. C. Jurs, *Appl. Spectrosc.*, **25**, 203 (1971).

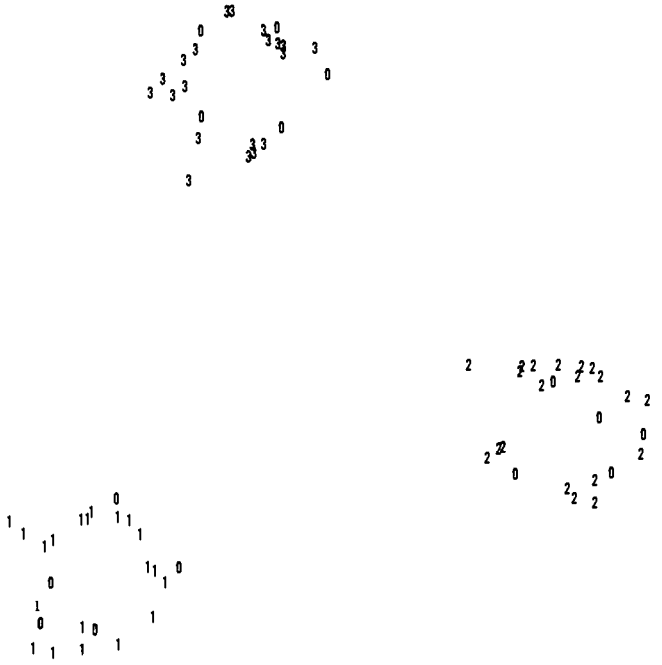

Figure 2. Nlm of three-class synthetic data from three-space to two-space.

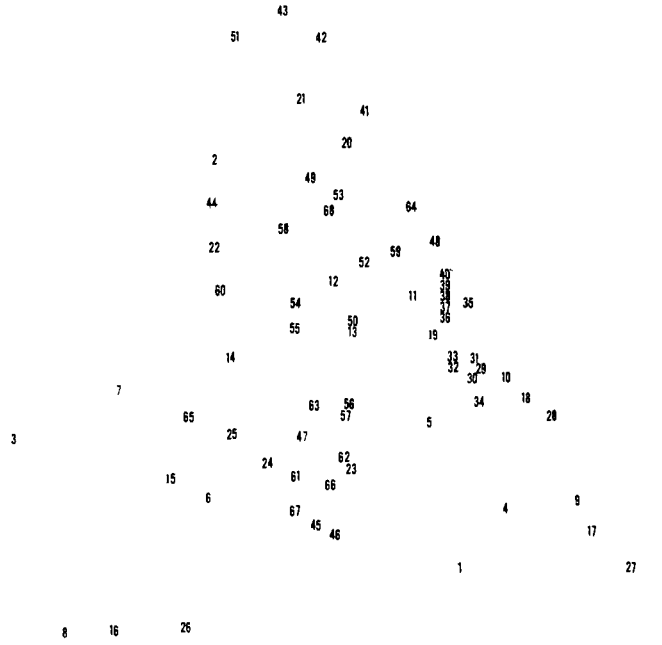

Figure 3. Nlm of acid–base data from six-space to two-space.

At this point computer techniques should be used but carefully supervised by the scientist.

It is important that the scientist get some feeling for the structure (again using the geometrical concepts discussed earlier) of the data. Obviously he cannot interpret the features in n space ($n > 3$), but the computer can be used to reduce the data to a more familiar two- or three-dimensional space. Clearly, this reduction can be done only approximately. There are a number of ways of performing such a task but probably the best is nonlinear mapping[22] (nlm), a technique which seeks to conserve interpoint distances. Every

(22) J. W. Sammon, Jr., *IEEE Trans Comput.*, C-18, 401 (1969).

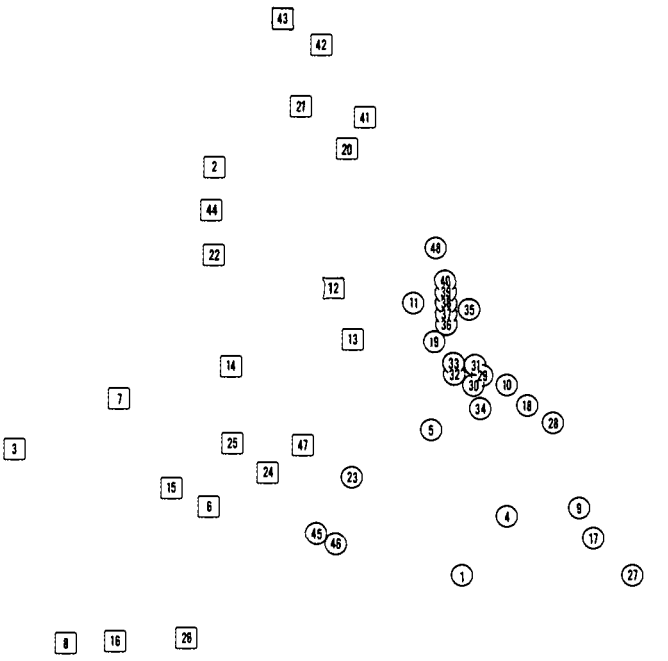

Figure 4. Acids (□) and bases (○) nlm from six-space to two-space.

point (i) in n space has a distance to every other point (j) defined as in eq 1. These distances can be calculated once and henceforth considered as constants, d_{ij}^*. The ideal reduction in two dimensions would have

$$d_{ij}^* = [(X_i - X_j)^2 + (Y_i - Y_j)^2]^{1/2} \equiv d_{ij} \qquad (9)$$

for all pairs (i,j). Since this cannot be done exactly, an error, E, is defined

$$E = \frac{1}{\sum_{i>j} d_{ij}^*} \sum_{i>j} \frac{(d_{ij}^* - d_{ij})^2}{d_{ij}^*} \qquad (10)$$

Examination of this function shows that it is nonlinear in 2 N unknowns (N = number of points), the unknowns being the (X,Y) coordinates of each point. Nlm is implemented by minimizing the error function using a nonlinear minimization method. This minimization is complicated but the resultant maps preserve the structure of data rather well making nlm an extremely valuable technique. For this paper, the minimization was done by the conjugate gradient method.[23]

Two additional and important points should be mentioned here. First, any distance measure can be used for d_{ij}^* and d_{ij} as long as it is monotonic and the derivatives of eq 10 exist. Second, the mapping can be from n space to m space where $n \geq m$. A two-dimensional map is most useful here but graphic display systems can easily handle three dimensions.

Figure 2 shows the results of mapping set 1 to two-space. This mapping was relatively simple because the original dimension was only three. The three distributions are well preserved in the mapping and the correct classification of unknown points (circles) would be easily done using only the map.

Figure 3 shows the nlm of set 2, where none of the points are classified. There is no obvious clustering, which is not unreasonable because there are various

(23) E. Polak, "Computational Methods in Optimization," Academic Press, New York, N. Y., 1971, p 53.

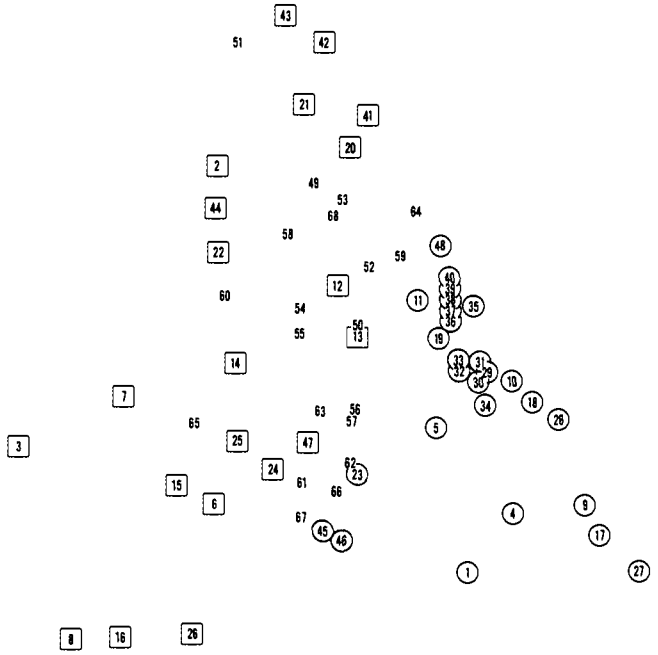

Figure 5. Acids (□), bases (○), and amphoterics nlm from six-space to two-space.

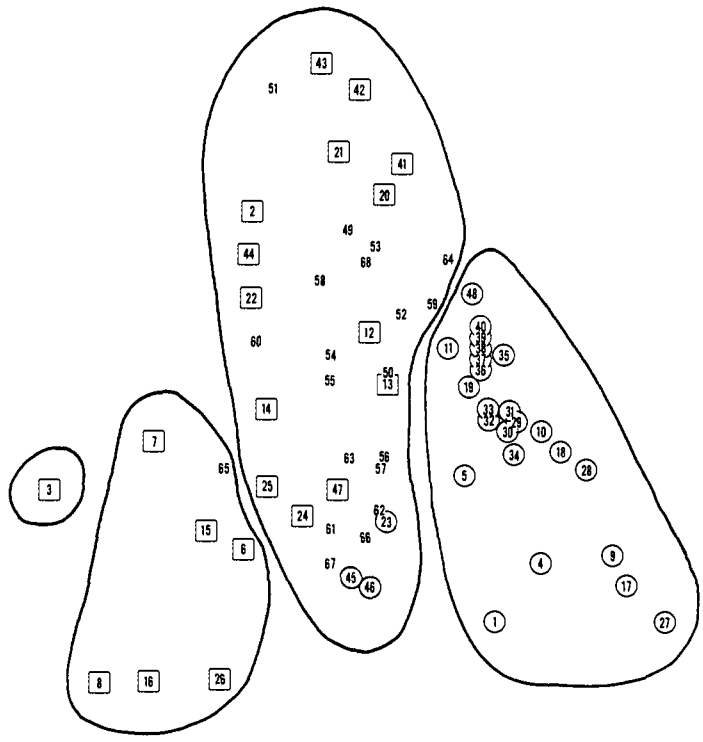

Figure 6. Four Q-mode clusters (□ = acid, ○ = base) using nlm as a display.

degrees of acidity and basicity. Set 2 is fundamentally very different from set 1 because set 1 contains distinct classes and set 2 does not. Figure 4 shows the same map as Figure 3 with only acids and bases shown. These two pseudoclasses are separated in Figure 4, whereas it is not possible to separate them using any one measurement. The map shows that the two classes can be separated using *indirect* measurements. Obviously, some of the measurements are more important than others, but it is the multivariant approach that allows separation.

There are excellent methods for elimination of redundancies among measurements and even ranking the measurements according to their relative importance, but they will not be discussed in this paper. Also, it is clear that an acid–base separation is not the only information in Figure 4. The variance perpendicular to the axis of separation is actually greater than the variance along the axis of separation. However, the significance of the perpendicular axis is not known. Figure 5 is the same as Figure 4 but includes the unknowns (amphoterics) to be classified. The reader can proceed to classify the unknowns, but should remember that the map is only an approximation of an n space; therefore, unknowns near the two-class interface should be classified by other methods.

Clustering

Methods of finding clusters in multidimensional data have been used for some time.[24,25] For this paper, a conceptually simple but effective clustering technique called hierarchical Q-mode clustering (hier) was used. Hier uses the similarity measurements from eq 2 in forming a similarity matrix. The matrix is scanned for the largest value and the two points producing this value are summed. Thenceforth, the two points are considered as one (a center of gravity) to calculate a new and smaller similarity matrix. This process is continued until all of the points are in one cluster. Center of gravity points, which represent more than one point in n space, are given more weight in order to reflect the size of the cluster represented by the center of gravity. The operator has three choices for obtaining clusters from hier. First, the desired number of clusters can be specified and the process will stop when a similarity measure produces a specified number of clusters. Second, a predetermined value can be set and when the similarity level reaches that value, the process stops and the clusters at that level are given. Third, the rate of change of the similarity level needed to produce the next clustering can be monitored and the process stopped when a change of, say, 5% occurs. This third method was used on the two data sets in this paper. Hier stopped at a similarity of 0.83 when applied to set 1. That is, the similarity started at 1.0 and changed slowly as new clusters formed. At 0.83, there were three clusters but the next step produced a large jump to 0.30, which is far greater than the specified 5% change. Further clustering was discontinued. It is not surprising that all 75 of the points fell correctly into one of the three classes. Again, set 1 is a trivial example, but if clustering does exist, even in much higher dimensional spaces, hier will find the correct clusters.

Using hier on set 2, again in the third mode as detailed above, gave the results shown in Figure 6. The first jump greater than 5% occurred when the similarity changed from 0.72 to 0.67. At 0.72, four clusters were formed. Figure 6 is the nlm map and is used instead of a table listing the elements in each cluster. It also shows a good "view" of the data structure because nlm preserves the global structure and hier preserves local structure. The two methods

(24) G. H. Ball, Data Analysis in the Social Sciences, Proceedings of the Fall Joint Computer Conference, Las Vegas, Nev., 1965.

(25) R. R. Sokal and P. H. Sneath, "Principles of Numerical Taxonomy," W. H. Freeman, San Francisco, Calif., 1963.

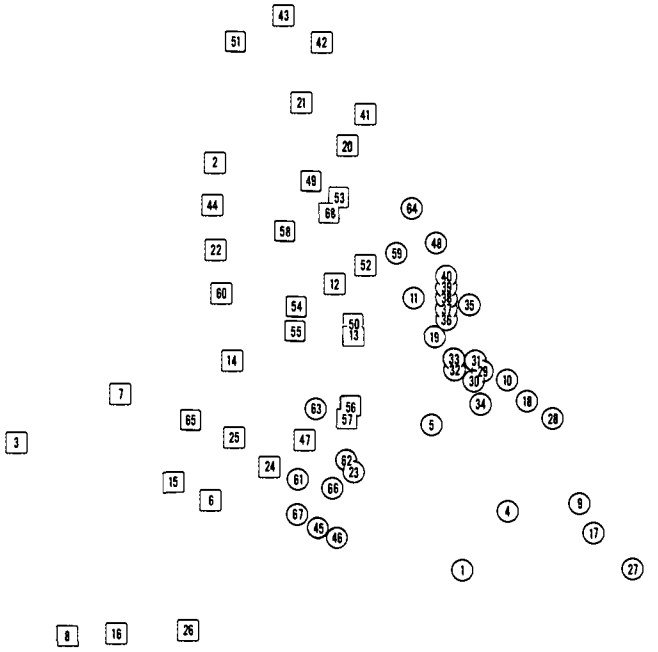

Figure 7. Final classification (□ = acid, ○ = base) using nlm as a display.

used together as in Figure 6 are excellent for unsupervised problems because they are complementary.

As expected, the clustering is not as distinct for set 2 as it is for set 1. Further clustering within the largest cluster, which is quite possible using hier, gave meaningless results because the clusters were not well separated. The three remaining clusters are reasonable when one considers their composition.

Classification

The most often encountered goal of a pattern recognition application is classification. Using a collection of knowns and a classification rule, a set of unknowns is classified. Mapping techniques such as nlm can be used for this purpose, with the scientist performing the actual classification. Clustering techniques can be useful in classifying unknowns by either of two methods. As one example, hier was applied to both the knowns and unknowns for set 1 and set 2. In this way, an unknown can be assigned to the class that holds a majority of knowns within the cluster containing the unknowns. A second method is to apply hier only to the training set. After the clusters are found, their centers of gravity are calculated. To classify an unknown, the cluster with the nearest center of gravity is found and the majority class is assigned to the unknown.

These methods of classification are merely follow-up procedures to unsupervised learning. There are, however, several classification methods[4,11,26,27] that are used strictly for supervised learning. These methods operate on the assumption that the classes are known and proceed to classify unknowns into one of the classes. One of these methods has been used extensively to analyze spectroscopic data. This method, a computerized learning machine,[11] employs a feedback procedure to find a hyperplane that divides the n space into two regions. If the two classes are linearly separable,[11] each region will contain one class. This method suffers from some unfortunate disadvantages, which are examined in a paper in preparation; therefore, it will not be discussed further here. As an alternative, the K-nearest-neighbor classification rule[28] (Knn) was used to classify the unknowns in set 1 and set 2. The characteristics of this method, which make it more desirable than the learning machine method, are beyond the scope of this paper. Suffice it to say that it is based on a firm statistical foundation and might possibly become the standard classification method by which all new and more sophisticated methods will be judged.

The K-nearest-neighbor rule is computationally and conceptually quite simple. An unknown is classified by the majority rule of the K-nearest knowns. In view of the relatively small data sets used in this paper and because a majority vote is easier when the number voting is odd, K selected for set 1 and set 2 was three. (At this point, note that all of the methods used in this paper work equally well for any number of classes.)

When Knn was applied to set 1, all of the unknown points were classified correctly. Again, this is not a great surprise. When Knn was applied to set 2, the unknowns (amphoterics) were classified as shown in Figure 7. The classification given to points 38, 57, and 63 clearly shows the danger of using only nlm for classification near the interface of two classes. (Local structure cannot be preserved exactly when mapping to a lower dimensional space.) It is rather difficult to "check" the results of the classifications made on the amphoteric oxides. Vanadium oxide is generally considered as being more acidic and lead oxide more basic but for many of the amphoterics, such as aluminum oxide, very little can be found on their preponderance as acids or bases.

Conclusion

The expressed purpose of this paper is to provide a broad scope introduction of pattern recognition as a tool to be used in making a direct connection between chemical data and desired results. It is hoped that through the use of examples, a flavor of what can be done will be given to the reader and that the broad applicability will be recognized. Pattern recognition should not be viewed as an attempt to remove the scientist from the data analysis part of experimentation. Nor should it be thought of as a black box within the computer that gives a machine a high degree of intelligence. Rather it is a combination of tools that can efficiently handle the tedious task of data reduction.

Most of the methods of pattern recognition are nicely suited to scientist supervision and interaction. Also, part of the power of pattern recognition to thoroughly extract information from a body of data is that the scientist can use *more than one* method for an application. In these cases, ultimate supervision is provided by the scientist who must decide whether or not the results "make sense" in the real world.

In describing the various branches of pattern recognition several important points have deliberately been avoided in order to keep the size of the paper

(26) Y. Ho and A. K. Agrawala, *IEEE Trans. Automat. Contr.*, AC-13, 676 (1968).
(27) G. Nagy, *Proc. IEEE*, **56**, 836 (1968).
(28) T. M. Cover and P. E. Hart, *IEEE Trans. Inform. Theory*, IT-13, 21 (1967).

manageable. Parametric methods were barely mentioned, the important preprocessing step was only lightly treated, and only one method for each approach was demonstrated. If sufficient interest is stimulated, these points will be treated in greater detail.

Acknowledgments. We express our thanks to J. W. Frazer, J. Harrar, and L. W. Hrubesh for carefully and constructively reviewing the manuscript. We would also like to acknowledge stimulating discussions with R. Ward and R. A. Anderson.

The Role of Ring Torsion in the Electrocyclic Transformation between Cyclobutene and Butadiene. A Theoretical Study

Kang Hsu,[1a] Robert J. Buenker,*[1a] and Sigrid D. Peyerimhoff[1b]

Contribution from the Department of Chemistry, University of Nebraska, Lincoln, Nebraska 68508, and Institute für physikalische Chemie, Johannes Gutenberg Universität, 65 Mainz, Germany.
Received November 8, 1971

Abstract: *Ab initio* SCF and CI calculations are reported which consider the effects of ring torsion (out-of-plane deformations) upon the reaction mechanism of the thermochemically induced electrocyclic transformation between cyclobutene and butadiene. It is found that conformations with both CH_2 groups perpendicular to the plane of the four carbons are strongly resistant to out-of-plane ring deformations but that structures with planar methylene groups are subject to a significant amount of such torsional displacements. The major effect of ring torsion upon the mechanism of this reaction is to decrease the CC distance R at which CH_2 rotation becomes favored relative to the corresponding value for a constrained reaction path in which torsion is not allowed; nevertheless the main conclusion of previous calculations is left unaltered, namely, that the rotational phase of this process occurs over a very narrow range of R. The calculations also indicate that formation of *trans*-butadiene, the ultimate product of the reaction, involves the cis isomer as an intermediate rather than direct conversion as a result of *simultaneous* CH_2 rotation and ring torsion.

In a previous communication[2] *ab initio* SCF and CI potential surfaces for the C_4H_6 isomers of cyclobutene and *cis*-butadiene were reported with the aim of providing detailed information about the minimum energy path followed by these systems in a thermochemically induced electrocyclic transformation. According to this study a partial opening of the cyclobutene ring occurs prior to any rotation of the methylene groups; at a certain separation of the carbon termini rotation becomes energetically likely and only after the complete rotation has occurred is further CC stretch to the *cis*-butadiene product favored.

The calculation of a reaction surface of this nature requires a sufficiently detailed examination of the energy dependence of *each* of the geometrical parameters. Because of the large number of these quantities, however, it becomes a matter of practical necessity, regardless of the method of calculation employed, to forego the complete optimization of each of these parameters and rather to assume certain fixed relationships for some of those species which do not appear to play a critical role in the process as a whole. Thus optimal values for the CH bond lengths and the HCH angles have been assumed in I, and C_2 or C_s symmetry of the nuclear framework is maintained throughout. With these assumptions the geometry search was then carried out in a four-dimensional space spanned respectively by the CC terminal bond distance R (see Figure 1), the CH_2 rotation angle θ (planar CH_2 groups, $\theta = 0°$; perpendicular, $\theta = 90°$), the CH_2 flapping angle α, and a fourth parameter γ ascribing a fixed relationship between the two distinct internal CC bond distances (double and single bonds in the end products).

Perhaps the most tenuous of the aforementioned assumptions is that at no point in the electrocyclic reaction between cyclobutene and *cis*-butadiene does out-of-plane deformation of the carbon ring, as described by an angle φ (planar ring, $\varphi = 0°$), occur. The semiempirical valence bond calculations of van der Lugt and Oosterhoff[3] for the same systems, for example, have predicted that torsion of the ring does play an important role and *optimal* values of φ as high as 40° are calculated for small values of R (close to the cyclobutene structure). More recently *ab initio* calculations of Radom and Pople[4] and of Dumbacher[5] have indicated that *cis*-butadiene itself may favor as much as 20° torsion relative to the planar conformation. Furthermore, the results of van der Lugt and Oosterhoff indicate that torsion can have a major effect on the transformation mechanism itself. It therefore seems necessary to investigate the effects of out-of-plane ring deformation in the framework of *ab initio* SCF and CI calculations similar to those described in I to determine whether the proposed mechanism is affected by such considerations. In addition, explicit attention will be given to the HCC angle β since exploratory calculations have indicated that this parameter

(1) (a) University of Nebraska; (b) Johannes Gutenberg Universität.

(2) K. Hsu, R. J. Buenker, and S. D. Peyerimhoff, *J. Amer. Chem. Soc.*, **93**, 2117 (1971); hereafter referred to as I.

(3) W. Th. A. M. van der Lugt and L. J. Oosterhoff, *ibid.*, **91**, 6042 (1969); also *Chem. Commun.*, 1235 (1968).

(4) L. Radom and J. A. Pople, *J. Amer. Chem. Soc.*, **92**, 4786 (1970).

(5) B. Dumbacher, Ph.D. Thesis, Mainz, June 1970; also see *Theor. Chim. Acta*, **23**, 346 (1972).