Spectral classification using pattern-recognition techniques. I. Feasibility with hydrogen as a model system

Keith L. Peterson and M. L. Parsons
Department of Chemistry, Arizona State University, Tempe, Arizona 85281
(Received 19 October 1976)

Pattern-recognition techniques are applied to the classification of atomic spectral transitions using hydrogen as a model system. Certain of the transitions have been randomly selected and treated as unclassified lines. Training of the known data at the 100% level has been achieved. Prediction of n_1 is 95.8%, while predictions of transitions of the type n_1, $l_1 \rightarrow l_2$ and n_1, $l_1 \rightarrow l_2$, $J_1 \rightarrow J_2$, are 83.4% and 66.67%, respectively. Using similar techniques, two unknown transitions in neutral atomic sodium have been classified.

I. INTRODUCTION

The classification of atomic spectra is primarily accomplished by searching for regularities in energy-level differences. Typically, this involves searching for constant intervals among transitions of two low levels (i.e., high wavelength) and transitions of one high and one low level (i.e., low wavelength). Low-wavelength lines should combine with low levels, and by initiating searches for these low levels, it is often possible to determine the lowest levels of the atom in question. Using these levels as a basis for further searches produces more levels. Relative intensities of transitions, isotope shifts, and Zeeman-effect data are all utilized in attempting to assign configurations and/or terms to the levels of the transitions in question. These classifications are also assisted by the use of various theoretical calculations such as Hartree-Fock methods. This combination of techniques is in many cases sufficient to obtain a fairly complete classification of spectra. However, one notes that of all observed lines which are listed in, for example, the NBS tables[1] only about 85% are classified, and many of these classifications must be considered tentative. In view of this, it is important to utilize new procedures to assist in the classification scheme based on pattern-recognition techniques.

This is a new and unique application of pattern recognition. The feasibility of the process will be demonstrated using atomic hydrogen as a model system. Later, it will be applied to atomic sodium in an attempt to classify two unknown transitions.

Pattern recognition has been applied successfully to a variety of chemical problems including structure-reactivity prediction as well as generation and prediction of mass spectra.[2,3]

Pattern recognition is a form of artificial intelligence which is capable of aiding the scientist in making a systematic analysis of multidimensional data. Thus, an interaction is possible between the investigator, who can supply intuition and intelligence, and the computer, which by utilizing pattern-recognition techniques can recognize relationships between data in a multidimensional space. A general statement of the pattern-recognition problem is: "Given a set of objects and a set of measurements made on those objects is it possible to find and/or predict a property of the objects that is not measurable but is known to be related to the measurements via some unknown relationship?"[4] A property is the information implicit in the data set in question.

In our work each type of atomic transition is designated as a category. For example, the transitions $3s$-n_2p $\frac{1}{2}$-$\frac{3}{2}$, $3s$-n_2p $\frac{1}{2}$-$\frac{1}{2}$, and $4p$-n_2d $\frac{3}{2}$-$\frac{5}{2}$ would be considered three different categories and identified as categories 1, 2, and 3. A given atomic transition can be represented by its wavelenth λ, energy-level difference ΔE, relative intensity RI, and two sets of quantum numbers n, l, and J, corresponding to the initial and final states of the transition. Thus, each transition can be thought of as a vector in nine-dimensional space, each of the above parameters being a coordinate axis. In pattern-recognition terminology each parameter is called a feature. Each transition, as represented by a nine-dimensional vector, is called a pattern.

To predict the classification of an unknown transition, lines with known features are used to "train" the computer to recognize the categories represented by the classified transitions. These classified lines are referred to as the training set. The unclassified transitions represented by the same known features are then assigned to a category (i.e., classification) on the basis of the training results. For the hydrogen example, an unclassified line has only three known features, λ, ΔE, and RI. Thus, the training set must be represented by only these three features. As more features are obtained, they may be used in the training set.

We have used a collection of pattern-recognition

techniques in a computer package obtained from Kowalski.[5] These techniques have been reviewed in detail elsewhere;[4,6-8] however, brief descriptions of the methods available in this package are included in Appendix A.

For hydrogen, data were included only for those lines with $n_1 = 1, 2, 3, 4, 5$.[22] For each n_1, n_2 was taken up to either 19 or 20 since this included all observed transitions. Lines are also observed with $n_1 = 6$, 7, etc.; however, this cutoff was used because it provided a large data set and because it includes all representative transitions. After $n_1 = 5$ all observed transitions are unresolved doublets of the type in which $l_1 = n_1 - 1$. For example, for $n_1 = 6$ the only observed transitions are $6h$-n_2i.

Energy-level differences were calculated directly from wavelengths and are, therefore, consistent with wavelengths although no corrections for air wavelengths to vacuum wavelengths were applied in this study. During the classification process the training data are immediately split up into two groups in accordance with the initial pattern-recognition procedure. One group contains those transitions for which $n_1 = 1,2$, whose wavelengths are all observed in the vacuum UV, and one contains those transitions for which $n_1 = 3,4,5$, whose wavelengths are in the UV and visible region. Thus, for almost the entire process, vacuum wavelengths are separated from air wavelengths. In either case ΔE is consistent with the wavelengths as measured.

Because of the differences in measurement techniques in these spectral regions, a consistent set of relative line intensities for hydrogen was not available. Therefore, we have used the quantity $(g_k)(A_{ki})$, where g_k is the statistical weight of the upper state, and A_{ki} is the transition probability of emission from upper state k to lower state i. The relative line intensity is a function of this quantity. A_{ki} values were taken from the NBS tables.[9] No A_{ki} values are given for fine-structure transitions. To obtain estimates for these transitions a table of line factors was used.[10] Thus, for example, the $1s$-$2p$ transition has $(g_k)(A_{ki}) = (6)(6.625 \times 10^8) = 37.590 \times 10^8$. From the line-factor table the transition $J = \frac{1}{2}$ to $J = \frac{3}{2}$ should be about twice as intense as the transition $J = \frac{1}{2}$ to $J = \frac{1}{2}$. Therefore, the $\frac{1}{2}$-$\frac{3}{2}$ transition will have $(g_k)(A_{ki}) = (\frac{2}{3})(37.590 \times 10^8) = 25.060 \times 10^8$, and the $\frac{1}{2}$-$\frac{1}{2}$ transition will have $(g_k)(A_{ki}) = (\frac{1}{3})(37.590 \times 10^8) = 12.530 \times 10^8$. The sum of the two numbers is of course equal to 37.590×10^8.

For unresolved doublets in the training data J_1 and J_2 were taken to be those J values corresponding to the component of the multiplet which theoretically should have the largest transition probability. For the unresolved doublet $5g$-$6h$, $J_1 = \frac{9}{2}$ and $J_2 = \frac{11}{2}$.

An initial experiment was performed in order to demonstrate the potential for the pattern-recognition approach. Using nine features the data[22] were investigated by the "minimal spanning tree technique." It was found that each category obtained by this unsupervised approach contained a unique type of transition; that is, each transition in a given category had a unique n_1, l_1, l_2, J_1, and J_2. Using these categories it was possible to separate the data using the multihyperplane separation technique (i.e., 100% training). Furthermore, it was found that when the data were categorized according to n_1 (5 categories, each category having a unique n_1), the data could be separated using four features: λ, ΔE, gA, and n_1. (See Appendix B for justification of the use of λ and ΔE.) The same also held true when the data were categorized according to n_1, $l_1 \rightarrow l_2$ (each category having a unique n_1, and l_1, l_2 pair), and the six features, $\lambda, \Delta E, gA, n_1, l_1, l_2$ used. Separation was also achieved for categories according to n_1, $l_1 \rightarrow l_2$, $J_1 \rightarrow J_2$ with nine features used.

Therefore, each type or category of transition was separable from all other categories, if the features defining these categories were included. This indicated a strong potential for pattern recognition in classifying transitions.

For an unknown transition in this model system, three features will be known: λ, ΔE, and gA. The general procedure for the classification process is to train the computer to predict n_1 from these three features; train and predict $l_1 \rightarrow l_2$ from the four features, λ, ΔE, gA, and n_1; and finally, train and predict $J_1 \rightarrow J_2$ using λ, ΔE, gA, n_1, l_1, and l_2. The test/prediction data set[22] was selected using a random number generator. A total of 24 transitions out of an original 130 were used as unknowns. There were 12 transitions originating from $n_1 = 1,2$; and 12 transitions originating from $n_1 = 3, 4, 5$. This leaves 106 transitions in the training set.

II. CLASSIFICATION PROCEDURE FOR HYDROGEN

An outline of the classification procedure is given in Table I. Several points merit further explanation. The training and test-set data are immediately split into two groups: one containing transitions originating from $n_1 = 1, 2$; the other containing transitions originating from $n_1 = 3, 4, 5$. This second test group is not linearly separable when categorized according to n_1. As indicated in parts I(C) and I(D) of Table I, we have utilized the Karhunen-Loeve transformation in order to effect the separation. The first two of the Karhunen-Loeve transform features contain approximately 93% of the total information or variance originally present in

TABLE I. Outline of hydrogen classification procedure.

I. Prediction of n_1

A. Categorize training data according to n_1. Run multi- and binary-hyperplane separation of data with three features (λ, gA, ΔE) to determine separability of categories.

B. Remove those categories which are separable from all other categories. Also remove any and all test patterns (data used as unknowns) which are predicted to lie in these categories. End result is two data sets: (1) categories which are separable from all other categories, (2) all remaining data. Data in (1) are already trained and classified (i.e. predicted) according to $n_1 = 1, 2$. Data in (2) contain all $n_1 = 3-5$ data.

C. All $n_1 = 3-5$ data are treated as follows. Obtain a Karhunen-Loeve transformation of the three-feature data (λ, gA, ΔE). Eliminate the third transformed feature for ease of visualization and plot the data using first and second transformed features as axes.

D. Define new categories such that each category has a unique n_1 (each n_1 does not necessarily have a unique category) and such that they will be linearly separable.

E. Run multi- and binary-hyperplane separations on data set to train and then classify (predict) unknown transitions according to n_1. All unknown n_1's are now predicted.

II. Prediction of $l_1 \rightarrow l_2$

A. Treat the two data sets from part I(B) in the manner described below. Treat them separately.

B. Redefine categories according to n_1, $l_1 \rightarrow l_2$. For example, $2s$-p is distinct from $2p$-s, $2p$-d, and $3s$-p.

C. Add n_1 as a fourth feature. The four features are now λ, gA, ΔE, n_1. For training data n_1 is known. For the test-prediction patterns n_1 will be that predicted in part I.

D. Run multi- and binary-hyperplane separations to train and predict $l_1 \rightarrow l_2$.

III. Prediction of J_1 and J_2

A. Treat the two data sets from part I(B) in the manner described below. Treat them separately.

B. Redefine categories according to n_1, $l_1 \rightarrow l_2$, $J_1 \rightarrow J_2$.

C. Add l_1 and l_2 as fifth and sixth features. The six features are now λ, gA, ΔE, n_1, l_1, l_2. For training data, l_1 and l_2 are known. For the test-prediction patterns, l_1 and l_2 will be values predicted in part II.

D. Run multi- and binary-hyperplane separations to train and predict $J_1 \rightarrow J_2$. All data are autoscaled before any other mathematical operations are performed.

the three features, λ, ΔE, and gA. Therefore, a two-dimensional representation of the data is a very good approximation to the original training set. This makes it visually convenient to define new categories so that the $n_1 = 3, 4, 5$ transitions will become linearly separable. The new categories are shown in Fig. 1. By using these new categories, it is possible to train 100% the $n_1 = 3, 4, 5$ transitions. Each category has a unique value of n_1, although each n_1 does not have a unique category. **Thus, prediction of any number of categories will still predict the correct n_1. In effect, we have utilized a stepwise linear classifier, in which the separating lines are approximations to curves of high order.**

The use of the multi- and binary-hyperplane separation technique to train and predict $J_1 \rightarrow J_2$ is a multistep procedure (See Table I, part III D). For $n_1 = 3, 4, 5$ the procedure is as follows. With the data categorized according to n_1, $l_1 \rightarrow l_2$, $J_1 \rightarrow J_2$, the transitions $3p$-n_2s $\frac{1}{2}$-$\frac{1}{2}$, cannot be separated from $3p$-n_2s $\frac{3}{2}$-$\frac{1}{2}$, and the transitions, $4p$-n_2s $\frac{1}{2}$-$\frac{1}{2}$, cannot be separated from $4p$-n_2s $\frac{3}{2}$-$\frac{1}{2}$. The two $3p$-n_2s categories are combined as are the two $4p$-n_2s categories, making the training data 100% separable. Prediction of J_2 can then be made since each category will have only one value of J_2. J_2 can then be used as a seventh feature (λ, ΔE, gA, n_1, l_1, l_2, J_2). The two previously combined categories are then split into their original categories (i.e., four categories, $3p$-n_2s $\frac{1}{2}$-$\frac{1}{2}$, $3p$-n_2s $\frac{3}{2}$-$\frac{1}{2}$, $4p$-n_2s $\frac{1}{2}$-$\frac{1}{2}$, and $4p$-n_2s $\frac{3}{2}$-$\frac{1}{2}$). With J_2 as a seventh feature, 100% training can be achieved and J_1 can be predicted.

For $n_1 = 1, 2$ transitions the same reasoning is

used. In this case the two categories $2p$-n_2s $\frac{1}{2}$-$\frac{1}{2}$ and $2p$-n_2s $\frac{3}{2}$-$\frac{1}{2}$ are combined as are the two categories $2s$-n_2p $\frac{1}{2}$-$\frac{3}{2}$ and $2s$-n_2p $\frac{1}{2}$-$\frac{1}{2}$. 100% training can be achieved with this categorization. Next, all $2p$-n_2s $\frac{1}{2}$-$\frac{1}{2}$ and $2p$-n_2s $\frac{3}{2}$-$\frac{1}{2}$ transitions are removed from the data set, so that J_1 is known. (All categories will have only one value of J_1 which is known from the first training run.) J_1 is then used as a seventh feature and J_2 is predicted. At this point the only J values which are not known are the J_1's belonging to $2p$-n_2s transitions. J_1 may be either $\frac{3}{2}$ or $\frac{1}{2}$ for this case and could be assigned on the basis of relative intensity or gA data. Alternatively, the $2p$-n_2s $\frac{3}{2}$-$\frac{1}{2}$, $2p$-n_2s $\frac{1}{2}$-$\frac{1}{2}$, $2s$-n_2p $\frac{1}{2}$-$\frac{3}{2}$, and $2s$-n_2p $\frac{1}{2}$-$\frac{1}{2}$ transitions may be removed from the data set, and the resulting two data sets trained to the 100% level.

Predictions and actual values are given in Table II, along with percentage correct for each step of the procedure. The n_2 term can be determined from the wavelength and originating level for each transition. By far the worst performance is on prediction of J values. It should be noted that in actual practice the assignment of J values could probably be done more accurately if a consistent set of relative-intensity data alone were used. Other potentially obtainable data such as observations of the Zeeman effect would further facilitate assignment of J's. Part of the problem for predicting J may be due to the ambiguity discussed previously in assigning J values to unresolved multiplets.

With the exception of J prediction, performance is quite good. In fact, it is probably better than most initial classifications of new atomic spectral systems which appear in the literature. Remember, however, that this method makes use of data obtained in the classical manner.

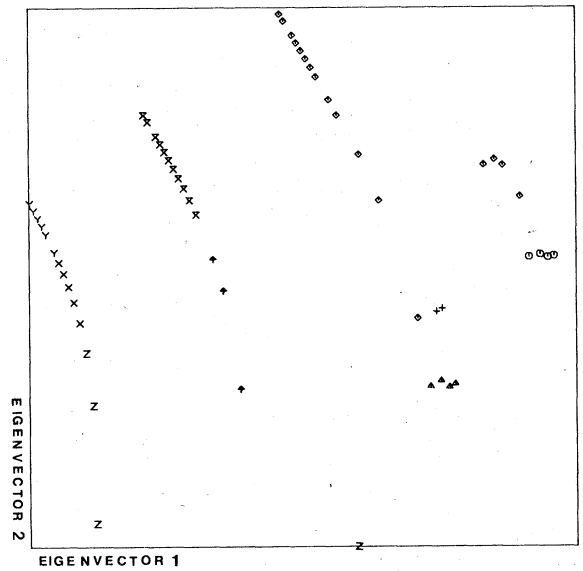

FIG. 1. Plot of the first two eigenvectors obtained from the Karhunen-Loeve transformation for the $n_1=3-5$ data set of hydrogen: 1 (○) and 5 (◇) correspond to $n_1=3$; 2 (△), 3(+), 6 (↑), and 7 (✕) correspond to $n_1=4$; 4(✕), 8(Z), and 9(Y) correspond to $n_1=5$. The original three features were λ, ΔE, and gA.

TABLE II. Summary of predictions for hydrogen.

(a) Correct predictions

$1s-6p$	$\frac{1}{2}-\frac{3}{2}$
$1s-10p$	$\frac{1}{2}-\frac{3}{2}$
$1s-13p$	$\frac{1}{2}-\frac{3}{2}$
$1s-15p$	$\frac{1}{2}-\frac{3}{2}$
$1s-17p$	$\frac{1}{2}-\frac{3}{2}$
$2s-4p$	$\frac{1}{2}-\frac{3}{2}$
$2p-6d$	$\frac{3}{2}-\frac{5}{2}$
$2p-9d$	$\frac{3}{2}-\frac{5}{2}$
$2p-16d$	$\frac{3}{2}-\frac{5}{2}$
$3s-5p$	$\frac{1}{2}-\frac{3}{2}$
$3d-7f$	$\frac{5}{2}-\frac{7}{2}$
$3d-10f$	$\frac{5}{2}-\frac{7}{2}$
$3d-17f$	$\frac{5}{2}-\frac{7}{2}$
$4f-8g$	$\frac{7}{2}-\frac{9}{2}$
$4f-18g$	$\frac{7}{2}-\frac{9}{2}$
$5g-15h$	$\frac{9}{2}-\frac{11}{3}$

(b) Incorrect predictions

True classification		Predicted classification	
$2s-3p$	$\frac{1}{2}-\frac{1}{2}$	$2s-n_2p$	$\frac{1}{2}-\frac{3}{2}$
$2p-3s$	$\frac{1}{2}-\frac{1}{2}$	$2p-n_2d$	$\frac{1}{2}-\frac{3}{2}$ or $2p$-n_2s $\frac{3}{2}-\frac{1}{2}$
$2p-4d$	$\frac{3}{2}-\frac{5}{2}$	$2p-n_2d$	$\frac{1}{2}-\frac{3}{2}$
$3s-5p$	$\frac{1}{2}-\frac{1}{2}$	$3s-n_2p$	$\frac{1}{2}-\frac{3}{2}$
$3s-6p$	$\frac{1}{2}-\frac{1}{2}$	$3p-n_2s$	$\frac{3}{2}-\frac{1}{2}$
$3p-5s$	$\frac{3}{2}-\frac{1}{2}$	$3p-n_2s$	$\frac{1}{2}-\frac{1}{2}$ or $3s-n_2p$ $\frac{1}{2}-\frac{3}{2}$
$4s-6p$	$\frac{1}{2}-\frac{3}{2}$	$5g-n_2h$	$\frac{9}{2}-\frac{11}{2}$
$4p-6s$	$\frac{3}{2}-\frac{1}{2}$	$4s-n_2p$	$\frac{1}{2}-\frac{1}{2}$

(c) Summary of percent correct

Quantum numbers predicted	Percent correct
n_1	95.8
$l_1 \rightarrow l_2$	83.4
$J_1 \rightarrow J_2$	66.7

III. APPLICATION TO NEUTRAL SODIUM

After a thorough investigation of the literature, it was determined that sodium would provide a suitable test element. It posesses two unclassified transitions and a suitable number of known transitions for training purposes. Assuming that the unknown transitions are outer-shell one-electron transitions, the energy-level differences indicate that the transitions must originate from either a $3s$ or $3p$ level. Any level greater than $3p$ would result in a final state having an energy greater than the ionization potential. Therefore, training data consisted only of transitions originating from the $3s$ and $3p$ levels. Self-consistent relative intensities for these lines were found in the MIT wavelength tables.[23] Figure 2 shows the training and test data plotted with the first two Karhunen-Loeve transformed features as axes. This is a reduction of the original three features, $\lambda, \Delta E$, and RI. Although n_1 is known, it is constant for all transitions and therefore has no classification value. The plots clearly show that transitions originating from a $3s$ level are separable from those originating from a $3p$ level. Furthermore, both unknowns fall into the $3p$ cluster. This is verified by multi- and binary-hyperplane separation. Unknown number 1 is very likely a $3p$-s or $3p$-d transition, while unknown number 2 is close to a forbidden transition.

The use of pattern-recognition techniques has provided us with a high probability of predicting the lower level for each of these transitions. Whereas further analysis by this technique was inclusive without additional features for training, this information is sufficient to allow us to predict these transitions. The $3p$ level must have either

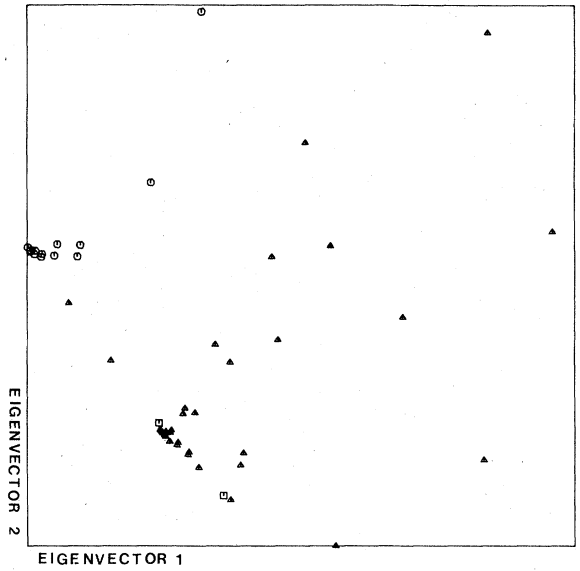

FIG. 2. Plot of the first two eigenvectors from the Karhunen-Loeve tranformation for Na. The original three features were λ, ΔE, and RI. 1 (Ⓓ) corresponds to a transition whose ground level is $3s$. 2 (Δ) corresponds to a transition whose ground level is $3p$.

$J=\frac{1}{2}$ or $J=\frac{3}{2}$. The energy-level difference may be calculated from the wavelength (corrected to vacuum)[11] and added to the energy of the two $3p$ levels. These data are summarized in Table III. The resulting final energy level ($3p\ \frac{1}{2}+\Delta E$) is closest to a $14s\ \frac{1}{2}$ state. Thus, the probable assignment is $3p$-$14s\ \frac{1}{2}$-$\frac{1}{2}$. In reality, this line is probably an unresolved doublet consisting of $3p$-$14s\ \frac{1}{2}$-$\frac{1}{2}$ and $3p$-$14s\ \frac{3}{2}$-$\frac{1}{2}$. The latter component of this doublet should

TABLE III. Summary of calculations for unclassified sodium transitions.

4198.3-Å line, $\Delta E=23\,812\ \text{cm}^{-1}$	5532.0-Å line, $\Delta E=18\,071\ \text{cm}^{-1}$
Possible lower levels	
$3p\ \frac{1}{2}$: $16\,956.183\ \text{cm}^{-1}$	$3p\ \frac{1}{2}$: $16\,956.183\ \text{cm}^{-1}$
$3p\ \frac{3}{2}$: $16\,973.379\ \text{cm}^{-1}$	$3p\ \frac{3}{2}$: $16\,973.379\ \text{cm}^{-1}$
Sum of lower level and ΔE	
$3p\ \frac{1}{2}+\Delta E=40\,769\ \text{cm}^{-1}$	$3p\ \frac{1}{2}+\Delta E=35\,028\ \text{cm}^{-1}$
$3p\ \frac{3}{2}+\Delta E=40\,786\ \text{cm}^{-1}$	$3p\ \frac{3}{2}+\Delta E=35\,045\ \text{cm}^{-1}$
Nearest energy levels	
$13p\ \frac{1}{2}, \frac{3}{2}$: $40\,705.68\ \text{cm}^{-1}$	$5p\ \frac{1}{2}$: $35\,040.27\ \text{cm}^{-1}$
$14s\ \frac{1}{2}$: $40\,769.5\ \text{cm}^{-1}$	$5p\ \frac{3}{2}$: $35\,042.79\ \text{cm}^{-1}$
$13d\ \frac{3}{2}, \frac{5}{2}$: $40\,798.8\ \text{cm}^{-1}$	
Probable assignment	
$3p$-$14s\ \frac{1}{2}$-$\frac{1}{2}$	$3p$-$5p\ \frac{3}{2}$-$\frac{3}{2}$

be the more intense.

In a similar manner, the final state of the second unknown is nearest to a $5p\frac{3}{2}$ level, although it is also very close to a $5p\frac{1}{2}$ level. Again, this line may be an unresolved doublet, the components of which have upper levels of $5p\frac{1}{2}$ and $5p\frac{3}{2}$.

We feel that with a second self-consistent set of relative intensities or other experimentally obtainable features, a classification procedure similar to the one used for hydrogen could be developed.

To test this idea, Landé g values were calculated for the lower and upper levels of all Na, $3s$ and $3p$ transitions. (We have assumed 100% LS coupling which should be a very good approximation for sodium.) Using the five features, λ, ΔE, RI, g_1, and g_2, the training data were linearly separable when categorized according to n_1, $l_1 \rightarrow l_2$, $J_1 \rightarrow J_2$. These five features are all potentially obtainable experimentally. The success of this experiment implies an even stronger potential for pattern-recognition techniques in classifying transitions.

IV. CONCLUSION

We have shown pattern recognition to be a useful tool in classifying atomic transitions. It is capable of high performance in predicting n and l values. Prediction of J's is not as good, but should be substantially improved by considering self-consistent sets of relative intensities from different excitation sources. Also, it is very often possible to determine J values experimentally utilizing Zeeman-effect data. We feel that training and prediction of all quantum numbers would be greatly improved if data which is potentially obtainable, but not always readily available from the literature, could be used. Such data includes the consistent relative-intensity and Zeeman-effect data mentioned above, as well as isotope shifts and perhaps even Stark-effect data. With this additional data, we could initiate the classification procedure using six or more features (assuming two or more self-consistent sets of relative intensities) which are all experimentally obtainable. These ideas are confirmed by our experiment with sodium.

As a logical extension of this work, we have applied pattern-recognition techniques to the classification of curium energy levels.[12] Used in conjunction with the conventional methods, pattern recognition has been shown to be a useful tool for atomic transition classification. Finally, we feel that pattern-recognition techniques are generally applicable to classification of spectral transitions, whether molecular or atomic.

ACKNOWLEDGMENTS

The authors wish to thank Professor B. R. Kowalski for supplying us with ARTHUR, the computer package of pattern-recognition techniques. We also wish to thank Don Anderson, who was responsible for converting ARTHUR to the computer system at Arizona State University. In addition, we thank the Computing Center at Arizona State University for computer time.

APPENDIX A: PATTERN-RECOGNITION TECHNIQUES

Although we have not made use of all the techniques available in the pattern-recognition package, the following concise summary of all the techniques are given for the sake of completeness.[5]

METHOD	DESCRIPTION
Preprocessing	
Change	Provides a variety of feature category and pattern alterations which provide flexibilty in data utilization.
Karhunen-Loeve transformation	Performs the Karhunen-Loeve transformation. The transformation is an orthogonal rotation of the n-dimensional coordinate system in question, such that the first new coordinate is in the direction of largest variance and thus contains the most separating information of the data set. The second coordinate chosen is orthogonal to the first and is in the direction of the second largest variance of the data set. This process is continued until n new coordinate axes have been chosen. The transformation affords a convenient means of feature reduction if the first several coordinates chosen contain most of the variance or information of the data set.[7]

METHOD	DESCRIPTION
Autoscale	Autoscales data. The process is similar to that used when one plots data on a two-dimensional graph. Such a process is necessary to avoid biasing results by data which may be expressed in small (or large) units and therefore be numerically quite large (small). The process transforms each feature to a mean of zero and a standard derivation of one without destroying separating information.[4]
Correlation-to-property selection	Produces weighted features which are linearly independent and ordered according to correlation-to-property weight, variance weight, or Fisher weight. It provides information for selecting the best features for separation.[4,13]
Tune	Generates all linear, quadratic and ratio combinations of original features.
Weighting	Evaluates the individual importance of each feature for the description of the property associated with the training-set patterns. Three weighting functions are available: correlation-to-property weight, variance weight, and Fisher weight.[4,13]
Utility and measurement analysis	
Correlation	Generates all feature-to-feature and feature-to-property correlations with their corresponding confidence intervals. Interfeature covariances are also given.[14,15]
Distance matrix	Calculates all interpattern distances. The following distance metrics are available: Mahalanobis, city block, and the ratio distance of Anders.[4,16]
Unsupervised learning	
Hierarchical clustering	Produces a dendrogram describing the hierarchical clustering (also known as Q-mode clustering) of the training-set patterns. The patterns are grouped at levels of similarity, where similarity is defined from the interpattern distances.
Minimal spanning tree	Generates a minimal spanning tree among all training-set patterns. A spanning tree is a connected graph containing all the training-set patterns and having no closed loops. A minimal spanning tree is a spanning tree whose total length is a minimum among all possible spanning trees. Clusters (or categories) are defined by breaking links between two points which are longer than specified. Any two adjacent broken links define a cluster.[8]
Supervised learning	
Bayes classification rule	Performs an approximate multivariate Bayes-rule classification. True probability distributions over each category for each feature are presumed to be unknown. Frequency histograms of the training set are used as approximation to the probability distributions. The routine is most suitable for very large data bases.[17]
K nearest neighbors	Predicts categories on the basis of K nearest neighbors, where K is 1, 3–10. A pattern belongs to that category which is represented most often among its K nearest neighbors.[18]
Least-squares regression	Performs a least-squares multilinear regression using all features, utilizing the generalized inverse method.

METHOD	DESCRIPTION
Multihyperplane separation	Iteratively develops $(n-1)$-dimensional hyperplanes separating given clusters (categories) of patterns, represented in an n-dimensional space. The routine uses negative-feedback training to develop the separating surfaces. This is referred to as a linear separator.[19]
Binary-hyperplane separation	Essentially the same as the multihyperplane separation method except that it is a binary classifier. It develops separating hyperplanes between two given categories at a time.[19,20]
Percentage nearest neighbors	Predicts categories on the basis of a given percentage of nearest neighbors. The routine is very similar to K nearest neighbors.
Stepwise multilinear regression	Performs a stepwise multilinear regression. Features used in the regression are determined by their contribution to the total variance of the data set.[21]
DISPLAY METHODS	
Nonlinear mapping	Performs a nonlinear mapping of the training set from its original n-dimensional space to two or three dimensional space. The routine minimizes an error function (which is a function of interpattern distances) in an attempt to preserve interpattern distances.[7]
Plotting	Provides line-printer or Calcomp plots of feature versus feature or feature versus category. It may be used to obtain plots of scaled data, Karhunen-Loeve transformation projections, weighted features or nonlinear mapping.

APPENDIX B

The use of both λ and ΔE as features may require some justification since they are universally thought to be redundant parameters. It is true that there is a mathematical expression relating the two, namely,

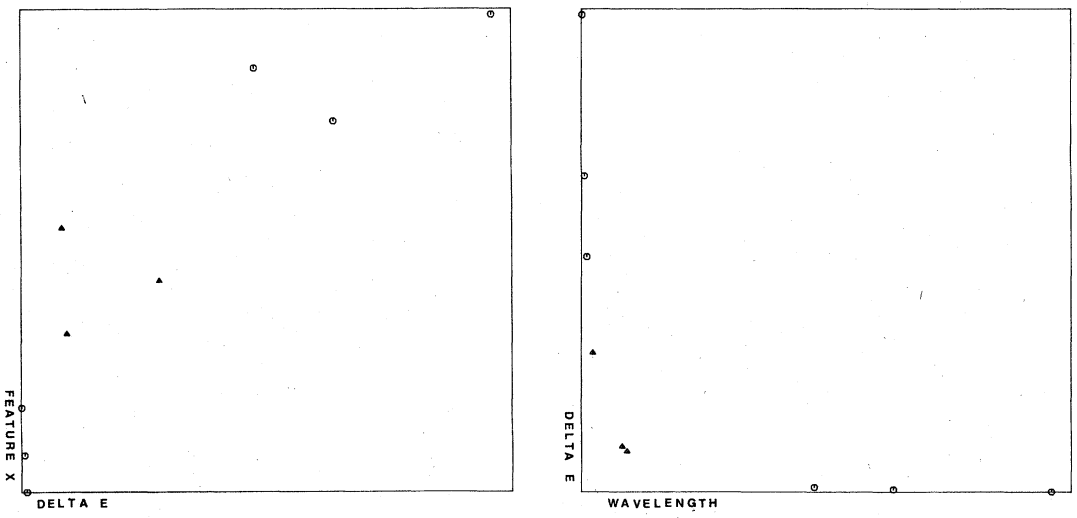

FIG. 3. Hypothetical data set showing that λ and ΔE are not redundant. In (a) a plot of feature x (where x may be any feature) versus ΔE shows category 1 (○) cannot be separated from category 2 (△) unless two separating lines are used. In (b) the data becomes separable when using both λ and ΔE.

$\Delta E = 1/\lambda$.

Obviously, this is an inverse relationship, and a plot of ΔE versus λ will be a hyperbola. All methods used in our pattern-recognition computer package are linear; that is, all separating surfaces are first-degree equations. In such a linear system, a $1/x$ relation cannot be redundant. This is borne out by the fact that the two features, λ and ΔE, have different correlation to category values. If λ and ΔE were related in a linear fashion, they would be redundant.

This argument may be better understood in geometrical terms. An idealized hypothetical example will be considered. Figure 3(a) is a plot of x versus ΔE where x may be any feature and may be considered analogous to RI used above. The points are denoted by ○ and △ corresponding to their respective (hypothetical) categories. It will be noted that no single linear separating surface can separate category one 100% from category two. Figure 3(b) is a plot of ΔE versus λ for the same points. Note that *on the basis of these two features alone it is possible to separate the two categories linearly*. By visualizing the data in three dimensions (ΔE, x, and λ as features), it is seen that the data are also linearly separable in this case. Thus, *the addition of λ as a third feature has added one dimension and made the data set* 100% *linearly separable*.

---

[1]W. F. Meggers, C. H. Corliss, and B. F. Scribner, *Tables of Spectral-Line Intensities*, 2nd ed., Natl. Bur. Stand. Monograph 145 (U.S. GPO, Washington, D. C., 1975).

[2]K. C. Chu, Anal. Chem. 46, 1183 (1974).

[3]J. Schechter and P. C. Jurs, Appl. Spectrosc. 27, 225 (1973).

[4]B. R. Kowalski and C. F. Bender, J. Am. Chem. Soc. 94, 5632 (1972).

[5]D. L. Duewer, J. R. Koskinen, and B. R. Kowalski, ARTHUR, available from B. R. Kowalski, Laboratory for Chemometrics, Department of Chemistry, BG-10, University of Washington, Seattle, Wash. 98195.

[6]B. R. Kowalski, in *Computers in Chemical and Biochemical Research* (Academic, New York, 1974), Vol. 2.

[7]B. R. Kowalski and C. F. Bender, J. Am. Chem. Soc. 95, 686 (1973).

[8]H. C. Andrews, *Introduction to Mathematical Techniques in Pattern Recognition* (Wiley-Interscience, New York, 1972).

[9]W. L. Wiese, M. W. Smith, and B. M. Glennon, *Atomic Transition Probabilities*, NSRDS-NBS 4 (U.S. GPO, Washington, D. C., 1966), Vol. 1.

[10]B. W. Shore and D. H. Menzel, *Principles of Atomic Spectra* (Wiley, New York, 1968), p. 449.

[11]B. Edlén, J. Opt. Soc. Am. 43, 339 (1953).

[12]K. L. Peterson, D. L. Anderson, and M. L. Parsons, following paper, Phys. Rev. A 17, 270 (1978).

[13]R. A. Fisher, Ann. Eugen. 7, 179 (1936).

[14]O. W. Davies and P. L. Goldsmith, *Statistical Methods for Research and Production* (Hafner, New York, 1972), p. 234.

[15]P. R. Bevington, *Data Reduction and Error Analysis for the Physical Sciences* (McGraw-Hill, New York, 1969), p. 123.

[16]O. U. Anders, Anal. Chem. 44, 1930 (1972).

[17]T. Y. Young and T. W. Calvort, *Classification, Estimation and Pattern Recognition* (Elsevier, New York, 1974).

[18]T. M. Cover and P. E. Hart, IEEE Trans. Inf. Theory IT-13, 21 (1967).

[19]N. J. Nilsson, *Learning Machines* (McGraw-Hill, New York, 1965).

[20]D. R. Preuss and P. C. Jurs, Anal. Chem. 46, 520 (1974).

[21]A. Ralston and H. S. Wolf, *Mathematical Methods for Digital Computers* (John-Wiley and Sons, New York, 1966).

[22]See AIP document No. PAPS PLRAA-17-261-4 for 4 pages of hydrogen spectral data. Order by PAPS number and journal reference from AIP, PAPS, 335 E. 45th N. Y., N. Y. 10017. Remit $1.50 for microfiche or $5 for photocopies.

[23]G. R. Harrison, *MIT Wavelength Tables* (Wiley, New York, 1939).