Development of a Novel Weighting Scheme for the *k*-Nearest-Neighbor Algorithm

R. A. FORBES, E. C. TEWS, and B. S. FREISER*

Department of Chemistry, Purdue University, West Lafayette, Indiana 47907

M. B. WISE

Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831

S. P. PERONE*

Lawrence Livermore National Laboratory, Livermore, California 94550

Received October 15, 1985

A weighting scheme for use with the *k*-nearest-neighbor pattern recognition algorithm that allows better class separation by precalculation of feature weightings is described. Two different feature selection methods have been tested with this weighting scheme by using chemical ionization (CI) mass spectrometry data obtained with transition metal ions as an ionization source.

INTRODUCTION

Pattern recognition has been applied to a wide variety of chemical problems, and numerous reviews on the subject have been published.[1-8] Some of its more recent applications include recognition of organic compounds by Fourier-transform infrared spectroscopy, interpretation of gas chromatography data, nuclear magnetic resonance spectral interpretation, and analysis of electrochemical systems.[9-19] Many applications of pattern recognition to mass spectral data have recently appeared in the literature. Electron-impact ionization is by far the most widely used ionization means and has been employed in studies ranging from the analysis of complex mixtures by gas chromatography/mass spectrometry to the recognition of steroids, and to the use of mass spectrometric data to predict the biological activity of antibiotics.[20-23] Pattern recognition has also been applied to the experimental optimization of field-desorption and fast atom bombardment mass spectrometry and for the location of homoconjugated triene and tetraene units in aliphatic compounds with NO chemical ionization.[24,25]

In our laboratory, the reactivities of laser-generated transition metal ions toward various types of compounds have been studied for several years.[26-30] A major goal of this work has been to evaluate the utility of metal ions as selective reagents for mass spectral identification of the functionality and structure of unknown compounds. In view of the potentially large data matrix generated from the reactions of different metal ions with various organic compounds, the application of pattern recognition techniques provides a particularly useful means for achieving these goals.

The *k*-nearest-neighbor algorithm (KNN) was chosen for several reasons. The algorithm does not assume that the data follow any underlying statistics. Conventional approaches that assume Gaussian statistics were considered inapplicable to the mass spectra data. The KNN classifier is easily applied to multiple categories, and it can easily handle multicentered clustering and class overlap.

This paper reports the development of a weighting scheme to be applied to feature selection, recognition, and prediction for this algorithm that improves class separation by the precalculation of feature weightings with information contained in the training set. Two feature selection algorithms were tested with these weightings: forward iterative feature selection with weighting optimization and successive subtraction of features with constant weightings.

These algorithms and the weighting scheme were tested on six data sets generated from metal ion CI mass spectra. Seventy two organics of six classes of compounds were reacted with Fe^+ to generate the first data set and with Y^+ for the second; the third consists of the combination of the data from the first two. Twenty four alkanes of three subclasses (linear, branched, and cyclic) reacted with Fe^+ and Y^+ comprise the fourth and fifth sets, and the last contains the combined data of the fourth and fifth sets.

METHODS

Data. The metal ion CI mass spectra for each of the compounds listed in Table I has been obtained as described above for both Fe^+ and Y^+. The data used for the recognition of the organic compounds consist of the branching ratios for the primary products generated by the initial reaction of the metal ion of interest with the organic neutral sample. Subsequent reactions of these product ions with the organic neutral sample (i.e., secondary, tertiary, etc. reactions) were not considered. Experimental conditions were adjusted such that predominantly primary products were observed. Under these conditions, the relative intensities of the products do not vary widely and have been found to be reproducible to better than 10%.

The branching ratios for each product ion are the ratios of the individual ion intensities (expressed as percent of base peak or actual magnitude of the signal voltage) to the sum of the intensities for all peaks. Other normalizations were studied, such as normalizing to percentage of base peak and geometrically normalizing each pattern vector to the same length, but significant changes in the recognition accuracies were not observed. It is possible that the weighting scheme reduces the effects of differing normalizations.

The mass numbers were adjusted to distinguish the products of Fe^+ from Y^+ in the following manner: first, the mass of the metal ion was subtracted from the ion mass (for ions corresponding to attachment of the metal ion to an organic fragment), and second, the number 1000 was added to the masses generated by yttrium. For example, the masses of $FeOH^+$ and YOH^+ would be adjusted to 17 and 1017 to be used as features by the pattern recognition programs.

Computer Software. All of the pattern recognition programs have been written for an IBM 9000 laboratory computer with the IBM version CS 9000 FORTRAN 77. The programs are stored on a 10-Mbyte hard disk, and back-up copies are stored on 1-Mbyte 8-in. diskettes. A hierarchical data management program has also been written for storage, retrieval, and other operations on the metal ion CI data. Along with the raw data, which are stored as sets of nominal ion mass and intensity, the name, formula, experimental conditions, and a reference for each compound are stored on 8-in. diskettes. Programs

0095-2338/86/1626-0093$01.50/0

Table I. Compounds Used for Recognition Experiments

Organics for Recognition of Six Classes					
alkane	alkene	ketone	aldehyde	ether	alcohol
butane	1-butene	butanone	propanal	ethyl ether	ethanol
pentane	1-pentene	2-pentanone	butanal	methyl butyl ether	1-propanol
hexane	1-hexene	2-hexanone	pentanal	ethyl propyl ether	2-propanol
heptane	(*E*)-3-hexene	3-heptanone	hexanal	propyl ether	1-butanol
2-methylpentane	3-methyl-1-butene	4-heptanone	heptanal	ethyl butyl ether	2-butanol
3-methylpentane	2-methyl-1-pentene	3-methyl-2-butanone	octanal	butyl ether	1-heptanol
2,3-dimethylbutane	4-methyl-1-pentene	3,3-dimethyl-2-butanone	2-methylbutanal	isopropyl ether	1-octanol
2,3-dimethylpentane	2,3-dimethyl-1-butene	2,4-dimethyl-3-pentanone	3-methylbutanal	methyl *tert*-butyl ether	2-methyl-2-propanol
cyclopentane	2,3-dimethyl-2-butene	cyclopentanone	2,2-dimethylpropanal	*sec*-butyl ether	2-methyl-2-butanol
1-methylcyclopentane	cyclopentene	methyl cyclopropyl ketone	benzaldehyde	ethylene oxide	2,2-dimethyl-1-propanol
cyclohexane	cyclohexene	3-methylcyclopentanone	cyclohexanecarboxaldehyde	propylene oxide	cyclopentanol
methylcyclohexane	vinylcyclohexane	cyclohexanone	cyclooctanecarboxaldehyde	tetrahydrofuran	cyclohexanol

Alkanes for Recognition of Three Subclasses		
linear	branched	cyclic
propane	methylpropane	cyclopropane
butane	3-methylpentane	cyclobutane
pentane	2,2-dimethylpropane	cyclopentane
heptane	2,3-dimethylbutane	methylcyclopentane
octane	2,3-dimethylpentane	cyclohexane
nonane	2,4-dimethylpentane	methylcyclohexane
decane	2,2,4-trimethylpentane	ethylcyclopentane
dodecane	2,2,3,3-tetramethylbutane	1,4-dimethylcyclohexane

have been written for calculation of weightings, unknown prediction, data set recognition, and feature selection.

All of the recognition and prediction programs employ the *k*-nearest-neighbor algorithm.[31] A classifier using this algorithm predicts the class of an unknown to be the same as that of the majority of its *k* nearest neighbors. For all of the experiments in this study, only the first nearest neighbor ($k = 1$) was used to effect classification, due to the relatively small size of the data sets, and the distance measure employed was the Euclidian distance in an *N*-dimensional feature space. Each compound is represented by a point in *N*-dimensional space (or pattern vector), and the distance from the origin along each feature axis is given by the intensity (or branch percentage) of that feature for a given compound.

The leave-one-out (LOO) algorithm[32,33] has been used to generate recognition accuracies for a given training set with a given set of features. Each pattern in the training set is treated as if it were unknown and classified by use of the rest of the training set. When each pattern has been classified in this manner, a total recognition accuracy for the training set is calculated. This recognition accuracy can then be used as a figure of merit for the classifier with that training set and set of features.

One of the goals of pattern recognition is to minimize the number of features required to effect class separation while maximizing the recognition accuracy through the elimination of features detrimental to class separation. Thus, an empirical feature selection algorithm is often used to map a classification problem down from the space of all features to a space of smaller dimensionality, which consists of only important, relevant features.

N. A. B. Gray showed that meaningless class separations could be obtained with linear discriminant functions if the ratio of patterns to features was less than or equal to 2.[34] Other workers have adopted a rule of thumb in pattern recognition studies; the ratio of patterns to features must exceed 3 in order to consider the separation of classes valid.[2] While no statistical proofs have been developed for the KNN classifier, empirical evidence suggests that the likelihood of producing a happenstance class separation increases as more features are selected. Thus, in our studies, the ratio of patterns to features is used as a qualitative index to the "goodness" of the class separation.

Weighting Schemes. A scaling factor or weighting of the features has been employed by other workers to improve class separations.[35] The multiplication of features by Fisher weightings or variability ratios has been found to improve classifications with discriminant functions as well.[2] The weighting scheme developed here is the extension of this concept to the multicategory KNN classifier.

For each calculation of distance between compounds, the class of one compound is known. The other compound is unknown (or treated as unknown in the LOO method). Thus, the a priori knowledge of the class of the known can be exploited to improve clustering of items of similar class and increase separations of items of unlike class. In this way, feature space can be scaled differently for each class of compounds in the data set, and in the comparison of the unknown to a known, the unknown is essentially projected into the particular space corresponding to the class of the known compound. The modified Euclidean distance calculation is given by

$$D_f(i,j)^2 = g_{f,\mathrm{cl}(j)}(I_i - I_j)^2$$

The distance square between unknown pattern *i* and known pattern *j* along feature axis *f* is given by the square of their intensity difference multiplied by a weighting factor for that feature and class of *j*, $g_{f,\mathrm{cl}(j)}$. Total distance squared from pattern *i* to *j* is given by the summation of each individual distance square over the features selected for the calculation.

Feature weightings are precalculated for a training set and stored in a disk file for use with feature selection programs. Several different types of weighting factors have been tried, such as the ratios of interclass variance to intraclass variance and the ratios of interclass variability to intraclass variability. A weighting scheme we have developed, which shows promise for mass spectral data, is the ratio of the total weighted average intensity for a feature to the intraclass weighted average intensity:

$$g_{f,\mathrm{cl}} = \frac{K(1/N)\sum_{i=1}^{N} I_{f,i}}{k(1/n)\sum_{j=1}^{n} I_{f,j}}$$

The weighted average of a particular feature is its average intensity (summation) multiplied by the number of occurrences of that feature (K for all compounds, and k within class cl). Since, for mass spectral data, the presence of a particular peak can be more important than its intensity, weighting by number of occurrences accounts for features that are common to a particular class. The weighting factors are then scaled by a factor of 100 for use in integer calculations, and if the denominator is 0, the weighting is set to an arbitrary maximum (10^4).

In this scheme the most important features are those that have the largest weighted averages within a particular class relative to their total weighted averages. These are the features that are unique to and numerous for a particular class. Since distance squares are weighted in the nearest-neighbor calculation, the weightings of a feature distinctive for a particular class are minima. For example, the square root of the weightings for feature axis 1016 (YO^+, attachment of oxygen to yttrium) are

feature	alkane	alkene	ketone	aldehyde	ether	alcohol
1016	100	100	4	5	6	6

No oxygen is present in alkanes or alkenes for the formation of YO^+, and that feature is never present in a spectrum for a compound of either class. YO^+, however, is the major peak for the other four classes of compounds that contain oxygen atoms. Thus, when a comparison along feature axis 1016 is made between a hydrocarbon and an "unknown", which has intensity for this feature, the distance is exaggerated by the large weighting factor. The unknown, therefore, appears further from that hydrocarbon and also all other hydrocarbons in the data set. Thus, the absence of 1016 for all hydrocarbons is important in their distinction from the other unlike classes, and this preinformation has been used to increase the Euclidean separation of the unlike classes.

Feature Selection Algorithms. While the only certain procedure for selecting the absolutely optimal subset of features is to perform an exhaustive search, this is not practical for most applications. Thus, a suboptimal approach is almost always used.

Two different heuristic feature selection algorithms have been examined in this study. The first algorithm that has been employed is selection of the best features by successive subtraction with total recognition accuracy as the selection criteria (SSTRA). Figure 1 provides a flowchart of this algorithm. With the execution of the program, the training set is first entered from a disk file, and then, the precalculated weighting array is entered from another disk file. The feature/weighting file contains the list of all features to be used in the search along with the weighting array (number of features by number of classes). These weightings remain constant throughout the feature selection process since no weighting optimization is performed with this first algorithm. The total recognition accuracy for the classification of the training set is obtained with these weightings and all features. Each feature, starting with the last, is then successively excluded, and the training set is reclassified without the feature. If the total recognition accuracy of the training set increases or remains unchanged with the exclusion of a feature, that feature is then permanently removed from the classifier. This process is repeated until all of the features have been tested in this manner.

The result of a feature search with SSTRA is a subset of features that is best, using the given weightings and search method applied. These features and their weighting array are then stored in another disk file and can be used by another program to predict the identity of a data set of "true" unknowns.

The next feature selection algorithm that has been employed

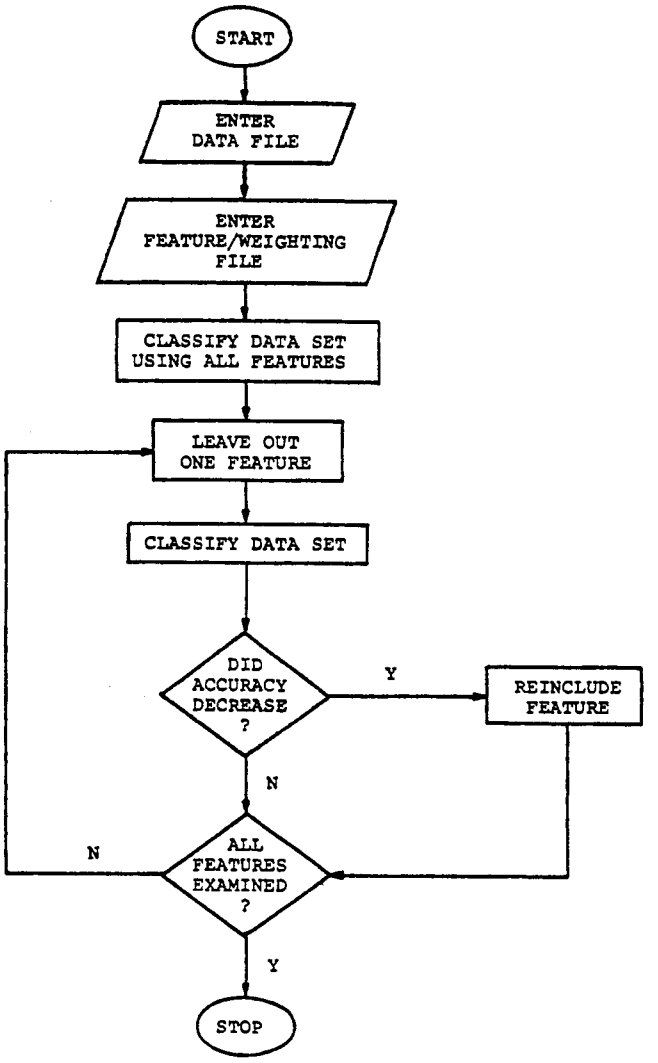

Figure 1. Flowchart for the successive subtraction algorithm SSTRA.

is the forward addition algorithm using the nearest-neighbor distance error criteria (FANNDE) diagrammed in Figure 2. The NNDE function replaces the discrete classification error with a related, continuous function for the purpose of performing a gradient search.[35] Distance calculations and unknown predictions are performed identically to the first program; however, a major difference from the first algorithm arises in that the weightings of the features selected by this second algorithm may be optimized at any point depending on the size of a projected error reduction term. Two modes of weighting optimization are available with this second program: in the first mode the weightings are optimized with respect to each feature and each class (multiclass weighting), and in the second the weightings are optimized only with respect to each other feature (identical weights for a feature over all classes—uniclass weighting).

As with the first feature selection program, the data file and the feature/weighting file are entered from disk storage, and the search is performed for the best subset of features from the list of features given in the feature/weighting file. The feature that can produce the minimum NNDE (related to the maximum recognition accuracy of the training set) when used singly is first selected. Each other feature is then individually combined with this best feature and an error term calculated. The feature that produces the minimum error when combined with the best feature is then added to the classifier. A similar search is performed for a third feature to combine with the first two and so on.

Before the inclusion of each feature, a gradient of the

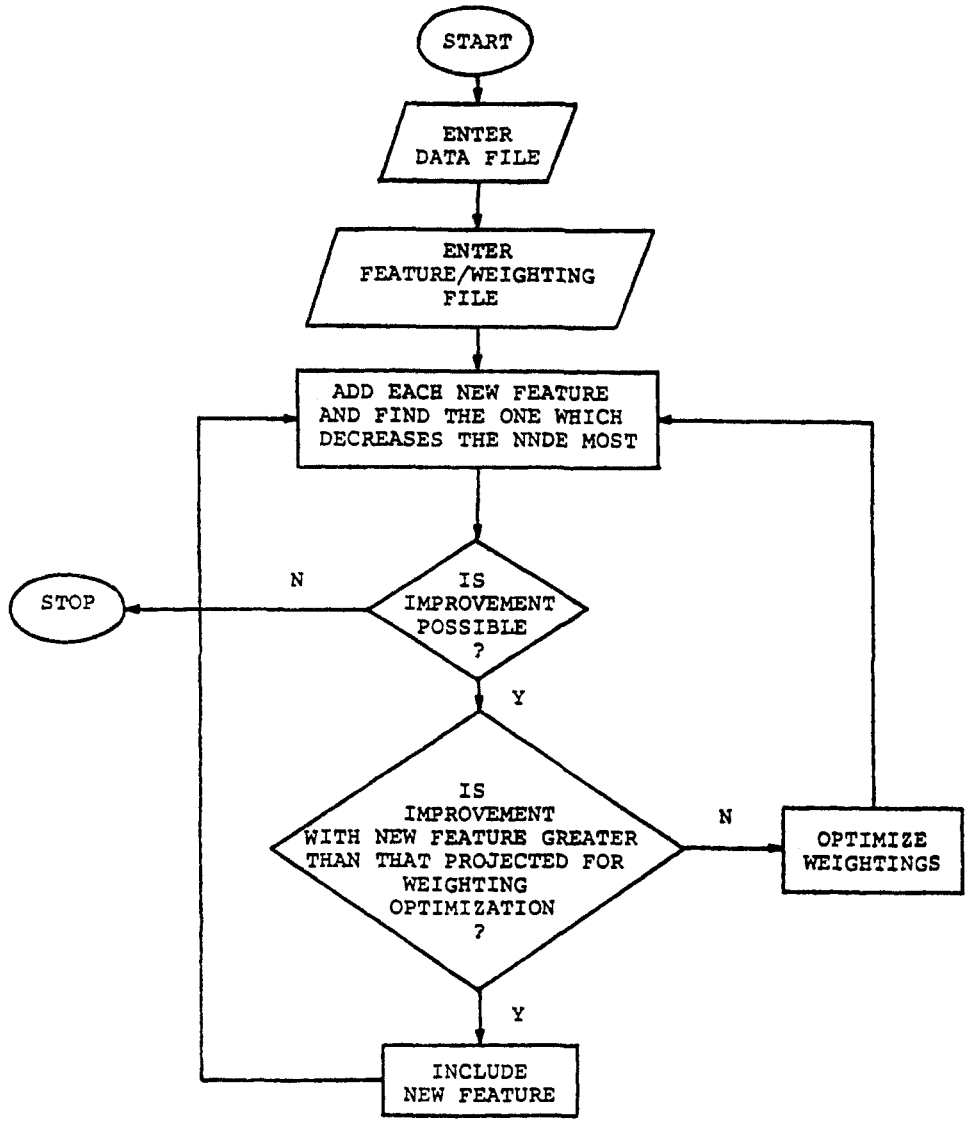

Figure 2. Flowchart for the forward addition algorithm FANNDE.

NNDE with respect to the weightings of the selected features is calculated along with a term that is the projected error reduction for weighting optimization. A comparison is made between the reduction of error produced by the inclusion of the new feature and the error reduction projected for optimizing the weightings of the features already selected. If this projected error reduction term is found to exceed the reduction in error for the inclusion of the new feature, the weightings for those existing features are then optimized with a gradient search algorithm. If the error can be reduced more by including the new feature, it is included in the model, and the selection process continues. The process of feature addition and weighting optimization continues until no further decrease in the error term can be obtained by either means. A more complete discussion of this algorithm with its error function has been described.[35]

Both heuristic searches are structured to converge to a minimum subset of features. A mass peak that is highly correlated with an already selected peak will not be included in the classifier, because it contains no new information and does not improve the performance of the classifier. There is a danger associated with using performance criteria in feature selection, however. It is possible for features to be chosen that separate the classes because of circumstantial rather than chemical information. This danger will increase with the size of the initial feature pool supplied to feature selection algorithms. In order to assess this danger, data sets were created with arbitrary class assignments, and the weighting/feature selection procedure was performed. Better than random classifications were obtained only when large subsets of features were selected, yielding unacceptable pattern-to-feature ratios. As with any supervised learning analysis, a separation of the classes is sought, and it is the job of the scientist to interpret the results and determine if this separation is meaningful.

Feature Searches Performed. Five different searches for an optimal subset of features have been performed. Two searches are performed with SSTRA; in the first, multiclass weighting is achieved by using the weighted average calculation, and in the second, an unweighted search is performed. Thus, a comparison can be made between the two procedures, and the effect of multiclass weighting can be assessed.

Three different searches are performed with FANNDE. Multiclass weighting optimization is performed starting from weighted averages and unit weightings, respectively, and in the third, unit weightings serve as a starting point for uniclass optimization. Thus, a comparison can be made between multiclass and uniclass weighting with optimization, and the effect of starting from different initial weightings can be determined.

The different searches are performed in order to approach the classification problem from several angles, since the results of heuristic searches are, by their nature, dependent on the path taken.

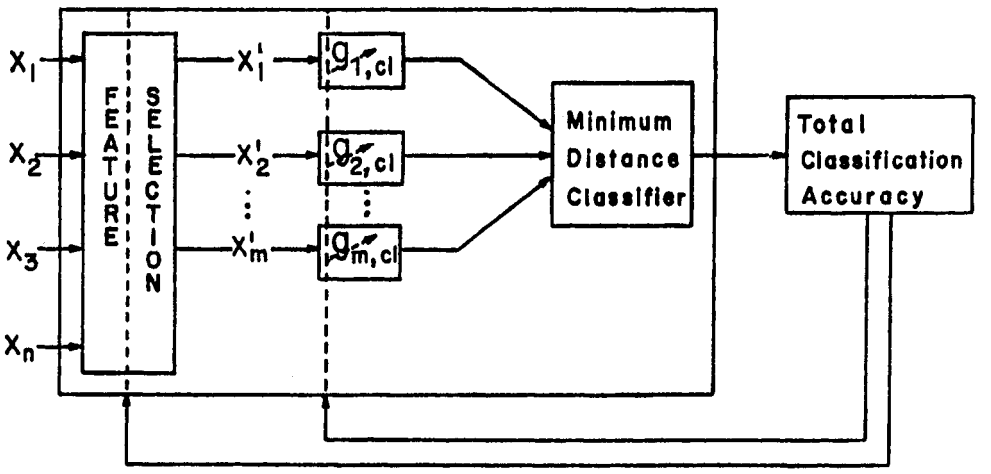

Figure 3. Training of the *k*NN classifier with feature weighting.

Table II. Recognition Accuracies for Six Organic Classes with Two Feature Selection Algorithms and Various Weighting Schemes

	SSTRA av wt	SSTRA unit wt	FANNDE av wt	FANNDE unit wt	FANNDE equal class wt
Recognition Accuracies Using Fe^+ Data					
alkane	75	92	75	92	92
alkene	67	25	50	42	42
ketone	33	33	25	25	25
aldehyde	50	42	50	0	0
ether	75	50	50	58	67
alcohol	92	58	92	100	100
total	65	50	57	53	54
P/F	4.2	9	10.3	7.2	7.2
Recognition Accuracies Using Y^+ Data					
alkane	83	67	92	92	67
alkene	50	42	50	42	50
ketone	92	100	83	100	83
aldehyde	75	50	50	50	50
ether	67	58	50	67	58
alcohol	100	83	100	83	83
total	78	67	71	72	65
P/F	5.1	4.8	12	7.2	8
Recognition Accuracies Using Both Fe^+ and Y^+ Data					
alkane	92	67	92	92	92
alkene	83	67	75	75	75
ketone	67	58	25	83	75
aldehyde	67	50	58	67	67
ether	75	64	58	42	58
alcohol	100	75	92	92	92
total	81	65	67	75	76
P/F	3.4	4.5	9	14.4	6

Table III. Recognition Accuracies for Three Alkane Subclasses Using Two Feature Selection Algorithms with Various Weighting Schemes

	SSTRA av wt	SSTRA unit wt	FANNDE av wt	FANNDE unit wt	FANNDE equal class wt
Recognition Accuracies Using Fe^+ Data					
linear	100	100	100	100	100
branched	100	63	100	100	100
cyclic	63	88	75	75	38
total	88	83	92	92	79
P/F	2.7	3	3.4	5.6	6
Recognition Accuracies Using Y^+ Data					
linear	88	88	100	88	88
branched	100	88	88	75	88
cyclic	75	100	88	100	100
total	88	92	92	88	92
P/F	3.4	1.6	3	4	3.4
Recognition Accuracies Using Both Fe^+ and Y^+ Data					
linear	100	100	100	100	100
branched	100	75	100	100	100
cyclic	88	75	88	50	50
total	96	83	96	83	83
P/F	3	3.4	4.8	4.8	4.8

RESULTS AND DISCUSSION

The results of the five searches performed with the data sets for the six organic classes and the data sets for the three alkane subclasses are presented in Tables II and III, respectively. With the first algorithm, SSTRA, the overall recognition accuracies for the three organic data sets are improved 15%, 11%, and 16%, respectively, for iron, yttrium, and the combination by using weighted averages vs. unit weightings. Clustering of the classes has, therefore, been improved by weighting each single feature differently for each class. A slight improvement in total recognition accuracy of the alkane subclasses is noted for two of the three data sets with weighted averages. No consistent improvement in the pattern-to-feature ratio is noted for either weighting scheme with this algorithm.

The effect of multiclass weighting vs. uniclass weighting on recognition with FANNDE is less clear-cut than with SSTRA. A slight improvement in recognition accuracy is noted for two of the three organic data sets for weighted averages vs. uniclass weighting and also for two of the three alkane data sets. A substantial improvement in the pattern-to-feature ratio is noted for the organic data sets with weighted averages, but no improvement is observed for the alkane data sets. No clear advantage is observed by starting with either weighted averages or unit weightings for multiclass optimization with FANNDE. Significant differences in recognition accuracy and/or the pattern-to-feature ratio are observed for the two procedures, however, unpredictably.

It would seem, therefore, that multiclass weighting can produce substantial improvement in recognition accuracies with SSTRA and minimize the number of features found with FANNDE. Variations in the total recognition accuracies of a few percent are probably not very significant for data sets of this small size. It appears, however, that each different search often converges to a different set of features and weightings, with very different individual class recognition accuracies as well as different total recognition accuracies. The explanation appears to be in the order with which each feature is examined by the search algorithm and the set of initial weightings assigned to the features.

Explaining Trial Variations. The interaction between weighting and feature selection is complex. Figure 3 diagrams the training of the one nearest-neighbor classifier (1NN) used in the recognition experiments reported here. Each pattern in the training set is entered as a vector $[X(i); i = 1, 2, ..., N]$, and a subset of these features $[X'(j); j = 1, 2, ..., M]$ is selected.

The subset vector is then multiplied by the weight vector for the selected features [$g(j,\mathrm{cl})$; $j = 1, 2, ..., M$], where the weight vector corresponds to the class to which the subset vector is being compared (cl = the class of the known pattern). The pattern is then classified by the minimum distance classifier. After the entire training set has been classified, a total recognition accuracy (or error term) is calculated as a figure of merit for the subset vector with the given weightings. This figure is compared to the accuracy of the last trial and the feature/weighting model is adjusted accordingly.

It is apparent that the initial weightings assigned to each feature as it is considered in the model can influence its inclusion or exclusion from the classifier as it is trained. Even a very good feature can be discarded if its initial weightings are poor. This fact is evident from the variations in accuracies and numbers of features selected for trials with different initial weightings.

Another important factor in explaining the variation in recognition accuracies with different weightings and searches is that the two algorithms used in this study are not optimal for finding the absolute maximum classification accuracies possible in all cases. The order in which the features are examined and excluded/included in the classifiers will influence the next feature that is chosen. While the only sure way to find the absolutely optimal subset of features is to perform an exhaustive search of all possible combinations of features with all possible weightings, convergence to a more consistent result is possible by improving the heuristic search.

The search for a better feature selection algorithm and weighting scheme is presently under study. A more intelligent algorithm, which is less subject to path constraints, is being developed. The results of the searches presented here are certainly encouraging.

CONCLUSIONS

Two different feature selection algorithms have been examined along with the effect of multiclass feature weighting. The classification accuracy for a data set is dramatically improved when a good subset of the set of all its features are selected for classification. For example, in the trial of SSTRA with the combined iron and yttrium data and using weighted averages, the total recognition accuracy improves from 68% to 81% while mapping feature space down from 74 features to 21 features. In general, SSTRA converges to a larger subset of features than FANNDE, often with higher accuracy. The forward addition algorithm can, however, sometimes achieve the same recognition accuracy for a data set with far fewer features and, thus, produce a more statistically valid class separation. Although a higher recognition accuracy is desirable, the statistical validity of the class separation must be considered. Multiclass weighted average feature weightings can improve recognition accuracies greatly with SSTRA, while the effect of using these precalculated weightings vs. unit weightings as a starting point for weighting optimization by FANNDE is not so dramatic. Multiclass weighting optimization can reduce the number of features required for a good class separation by FANNDE.

The speed of the analysis and the fact that the unknown does not have to be present in the data set make pattern recognition attractive for online applications. Some of the objectives of work in our laboratory will be to expand the technique to more complex molecules of interest, such as multiple functionalities, isomeric compounds, and biologically active samples, and the range of metal ions included in the study will be expanded. Another long-range goal is the application of artificial intelligence to the optimization of experimental parameters involved in the metal ion FTMS experiment such that the maximum information can be obtained.

ACKNOWLEDGMENT

We acknowledge the Division of Chemical Sciences in the Office of Basic Energy Sciences in the U.S. Department of Energy (DE-AC02-80ER10689) for supporting the metal ion chemical ionization research and to the National Science Foundation (CHE-8310039) for providing funds for the advancement of FTMS methodology. R.A.F. acknowledges the support of Lawrence Livermore National Laboratory and the Office of Naval Research. We also thank fellow research group members for their help in collecting the metal ion data: Denley Jacobson, William Stanton, Leo Lech, and Beth Stanko.

REFERENCES AND NOTES

(1) Jurs, P. C.; Isenhour, T. L. *Chemical Applications of Pattern Recognition*; Wiley: New York, 1975.
(2) Varmuza, K. *Pattern Recognition in Chemistry*; Springer-Verlag: New York, 1980.
(3) Veress, G. E. *TrAC, Trends Anal. Chem. (Pers. Ed.)* **1982**, *1*, 374.
(4) Habbema, J. D. F. *Anal. Chim. Acta* **1983**, *150*, 1.
(5) Hippe, Z. *Anal. Chim. Acta* **1983**, *150*, 11.
(6) Leegwater, D. C.; Leegwater, J. A. *TrAC, Trends Anal. Chem. (Pers. Ed.)* **1984**, *3*, 66.
(7) Ritter, G. L.; Lowry, S. R.; Wilkins, C. L.; Isenhour, T. L. *Anal. Chem.* **1975**, *47*, 1951.
(8) Kowalski, B. R.; Bender, C. F. *J. Am. Chem. Soc.* **1972**, *94*, 5632.
(9) Frankel, D. S. *Anal. Chem.* **1984**, *56*, 1011.
(10) Parrish, M. E.; Good, B. W.; Jeltema, M. A.; Hsu, F. S. *Anal. Chim. Acta* **1983**, *150*, 163.
(11) Dunn, W. J., III; Stalling, D. L.; Schwartz, T. R.; Hogan, J. W.; Petty, J. D.; Johansson, E.; Wold, S. *Anal. Chem.* **1984**, *56*, 1308.
(12) Jellum, E.; Bjornson, I.; Nesbakken, R.; Johansson, E.; Wold, S. *J. Chromatogr.* **1981**, *217*, 231.
(13) Kowalski, B. R.; Bender, C. F. *Anal. Chem.* **1972**, *44*, 1405.
(14) Bink, J. C. W. G.; Van't Klooster, H. A. *Anal. Chim. Acta* **1983**, *150*, 53.
(15) Thomas, Q. V.; Perone, S. P. *Anal. Chem.* **1977**, *49*, 1369.
(16) Burgard, D. R.; Perone, S. P. *Anal. Chem.* **1978**, *50*, 1366.
(17) Schachterle, S. D.; Perone, S. P. *Anal. Chem.* **1981**, *53*, 1672.
(18) Byers, W. A.; Perone, S. P. *Anal. Chem.* **1983**, *55*, 615.
(19) Byers, W. A.; Freiser, B. S.; Perone, S. P. *Anal. Chem.* **1983**, *55*, 620.
(20) Chien, M. *Anal. Chem.* **1985**, *57*, 348.
(21) Rotter, H.; Varmuza, K. *Anal. Chim. Acta* **1978**, *103*, 61.
(22) Brent, D. A.; Roth, B.; Johnson, R. L.; Brunner, T. R. *Biomed. Mass Spectrom.* **1981**, *8*, 440.
(23) Tsao, R.; Voorhees, K. J. *Anal. Chem.* **1984**, *56*, 368.
(24) Van Der Greef, J.; Tas, A. C.; Bouwman, J.; Ten Noever Brauw, M. C.; Schreurs, W. H. P. *Anal. Chim. Acta* **1983**, *150*, 45.
(25) Brauner, A.; Budzikiewicz, H.; Boland, W. *Org. Mass Spectrom.* **1982**, *17*, 161.
(26) Burnier, R. C.; Byrd, G. D.; Freiser, B. S. *J. Am. Chem. Soc.* **1981**, *103*, 4360.
(27) Byrd, G. D.; Burnier, R. C.; Freiser, B. S. *J. Am. Chem. Soc.* **1982**, *104*, 3565.
(28) Jacobson, D. B.; Freiser, B. S. *J. Am. Chem. Soc.* **1983**, *105*, 5197.
(29) Jacobson, D. B.; Freiser, B. S. *J. Am. Chem. Soc.* **1983**, *105*, 7484.
(30) Jacobson, D. B.; Freiser, B. S. *J. Am. Chem. Soc.* **1983**, *105*, 7492.
(31) Cover, T. M.; Hart, P. E. *IEEE Trans. Inf. Theory* **1967**, *IT-13*, 21.
(32) Thomas, Q. V.; DePalma, R. D.; Perone, S. P. *Anal. Chem.* **1977**, *49*, 1376.
(33) Fukunaga, K. *Introduction to Statistical Pattern Recognition*; Academic, New York, 1972.
(34) Gray, N. A. B. *Anal. Chem.* **1976**, *48*, 2265.
(35) Byers, W. A.; Perone, S. P. *Anal. Chem.* **1980**, *52*, 2173.