Mass Spectrometric Analysis for Distinction between Regular and Premium Motor Gasolines

Masatoshi ICHIKAWA†, Naomi NONAKA, Ichiro TAKADA and Sigeru ISHIMORI

Suzuki Motor Corporation Research & Development Center, Hamamatsu-Nishi, P.O. Box 1, Hamamatsu 432-91, Japan

A method to estimate the octane number of motor gasoline by mass spectrometry (MS) has been studied, by first examining whether the octane number can be estimated from its mass spectrum (MS). The MS of 29 different regular gasolines and 32 premium gasolines, sampled in the market from spring to autumn, were measured. We studied whether it is possible to extract any available parameters for clearly distinguishing between regular and premium gasolines, by applying both feature-selection and pattern-recognition methods to MS. It was found that a clear distinction between regular and premium gasolines could be made using MS information.

Keywords Mass spectrometry, pattern recognition method, octane number, motor gasoline

The motor gasoline marketed in Japan can be broadly classified into two types, regular and premium, which are claimed to have octane numbers of at least 89 and 96, respectively. Whether motor gasoline purchased in the market is regular or premium is basically determined by measuring its octane number using a research octane-number assessing method[1] prescribed by the American Society for Testing and Materials (ASTM). Unfortunately, since this method requires not only elaborate hardware, but also considerable skill, it cannot be readily employed by inexperienced personnel.

The various reported methods for octane-number estimation by applying regression analysis to data obtained by instrumental analysis include methods to estimate from proton NMR[2-4] and gas chromatography[5-7] studies. There is also the method of infrared (IR) absorption spectrometry using the near-infrared absorption spectrum (15150 – 8230 cm^{-1}).[8] Furthermore, we earlier applied a pattern-recognition method to proton NMR data, thus demonstrating the ability to visually estimate the octane number in a two-dimensional pattern space and to identify any differences in the composition.[9-11] Moreover, we showed that IR data (3200 – 2800 cm^{-1}) are available for estimating the octane number.[12]

In this study we intended to readily and quickly determine whether a given sample of motor gasoline obtained in the market is regular or premium. An attempt was thus made to distinguish between the two types by analyzing the mass spectra (MS) of gasoline using a supervised learning pattern-recognition method, an unsupervised learning pattern-recognition method as well as other related methods. The encouraging results of the study are presented below.

† To whom correspondence should be addressed.

Experimental

Samples

Twenty-nine samples of regular gasoline and 32 of premium gasoline were bought from service stations in the market over the period from spring to autumn.

Equipment and measurement conditions

The analytical conditions are given in Table 1. For the gas chromatograph-mass spectrometer (GC-MS) system, a GCMS-QP1000 (Shimadzu, Kyoto, Japan) was used. MS were measured from 30 to 200 m/z. For the column, a fused-silica capillary column (Chemical Inspection & Testing Institute, Tokyo, Japan) was used, the innerface of which was inactivated. Using this column, injected samples were sent into the mass spectrometer without separation. To introduce samples into

Table 1 Analytical conditions

Column	fused silica capillary column (inactivating chemical treatment)
Oven temp.	200°C
Injection temp.	230°C
Separator temp.	250°C
Ion source temp.	250°C
Ionization voltage	70 eV
Ionization current	60 μA
Split ratio	80:1
Sample	1 μl

the GC-MS, a split mode of a split/splitless inspection device SPL-G9 (Shimadzu, Kyoto, Japan) was used. The mass spectrum of a motor gasoline is shown in Fig. 1 as an example.

Interpretation of MS data

From the MS of each sample, nine fragment ions related to aromatic, naphthenic, olefinic and branched paraffinic compounds were extracted and the relative intensities of the fragment ions and ratios of two specific peaks were calculated. These values were used as parameters for a later analysis.

Table 2 gives the fragment ion's mass number as well as combinations of the fragment ion's mass numbers which were selected to analyze the octane number. The mass numbers and combinations of the chosen mass numbers are as listed in the footnote of Table 2. The arranged MS data for 61 kinds of motor gasoline are listed in Table 3.

Data analyses

The data interpreted as mentioned above were standardized against a mean of 0 and a variance of 1; the standardized data were then analyzed using a supervised learning pattern-recognition method, an unsupervised learning pattern-recognition method as well as other related methods.

For the supervised learning pattern-recognition method and related classifying methods, a discriminant analysis[13] and the k-nearest neighbor method (kNN)[14] were used.

For the unsupervised learning pattern-recognition method and related classifying methods, a principal-component analysis[15], and a minimal spanning tree (MST) method[16,17] were applied.

The discriminant analysis used involved a step-wise discriminant analysis in the BMDP (Biomedical Computer Programs-P) program package for multivariate analysis.[18] According to this method, variates that are useful for discrimination are selected by partial λ statistics; a linear discriminant function based on minimizing Mahalanobis' generalized distance is calculated from these variates while also developing a linear discriminant function to maximize the correlation ratio (canonical discriminant analysis).

For all other methods, the pattern-recognition program package ARTHUR was used. Before classification, useful features for class discrimination were selected by utilizing the Fisher ratio.[19]

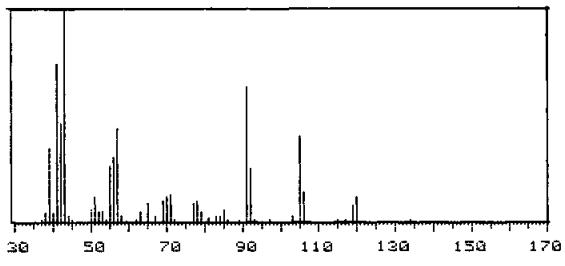

Fig. 1 Example of the MS spectrum of motor gasoline.

Table 2 Interpretation of the MS

Symbol	Mass number or combination m/z	Fragment ion
M1	43	$C_3H_7^+$
M2	57	$C_4H_9^+$
M3	71	$C_5H_{11}^+$
M4	91	$C_7H_7^+$
M5	105	$C_8H_9^+$
M6	119	$C_9H_{11}^+$
M7	120	$C_9H_{20}^{+\cdot}$
M8	41/M1	$C_3H_5^+$
M9	55/M2	$C_4H_7^+$

M1 - M3: information concerning the paraffin content and branching; M4 - M7: information concerning the aromatic compounds content; M8, M9: information concerning the naphthene, olefin content and branching.

Results and Discussion

Feature selection by SELECT method

Nine sets of MS data (M1 to M9) were subjected to feature selection by the SELECT method[20,21] in order to select those parameters valid for classification. In this study the Fisher ratios were used to evaluate the discriminatory power of the features. What were eventually selected were five sets of MS data (M4, M1, M2, M8 and M9), indicating that the main compositional difference between regular and premium gasoline largely involves the content of aromatic compounds and the degree of branching. All analyses, except for the discriminant analysis, were attempted by means of these parameters selected by the SELECT method.

Application of supervised learning pattern-recognition and related methods

Application of discriminant analysis. A step-wise discriminant analysis was applied to nine sets of MS data (M1 to M9). Four values (M4, M1, M2 and M8) were selected by this analysis as being useful for discrimination. The results of a discriminant analysis using the minimization of Mahalanobis' generalized distance and a canonical discriminant analysis demonstrate that regular and premium gasolines can be 100% discriminated from each other on the basis of the MS. The canonical variate of each sample is plotted on the canonical variate axis in Fig. 2. The equation of the discriminant variate used in this analysis was

$$Z = 0.08171 \times \mathrm{M1} - 0.06461 \times \mathrm{M2} - 0.09707 \times \mathrm{M4} + 0.03816 \times \mathrm{M8} + 0.48786. \quad (1)$$

The value of the canonical correlation coefficient was 0.964.

The center (0.175) between the mean of Group 1's canonical variates (3.73) and that of Group 2 (−3.38) can

Table 3 MS data of each sample

Sample	M1	M2	M3	M4	M5	M6	M7	M8	M9	Class
1	100.0	47.2	15.5	57.7	31.2	9.5	10.8	84.1	78.0	regular
2	78.8	49.5	14.1	100.0	38.0	7.4	11.0	79.1	55.6	premium
3	100.0	44.1	13.2	55.7	37.3	10.0	12.9	72.9	49.2	regular
4	78.8	47.7	13.7	100.0	40.0	7.5	12.4	89.6	71.1	premium
5	60.6	27.0	9.2	100.0	20.0	4.4	6.5	85.5	72.2	premium
6	100.0	38.3	9.5	33.6	2.6	0.7	0.0	67.7	30.5	regular
7	100.0	49.4	14.4	28.5	12.5	3.7	3.6	80.0	76.3	regular
8	95.5	63.5	17.3	100.0	22.4	1.8	4.1	68.2	27.8	premium
9	100.0	45.9	14.7	65.7	41.1	9.0	13.5	75.0	60.3	regular
10	91.2	73.2	17.9	100.0	22.2	5.7	8.3	71.9	24.5	premium
11	100.0	52.6	15.2	73.3	28.4	7.0	8.1	85.3	68.6	regular
12	52.1	27.5	8.1	100.0	30.2	3.2	8.9	99.0	80.4	premium
13	100.0	57.2	16.8	39.4	18.5	7.2	6.5	78.1	57.2	regular
14	73.4	45.7	14.0	100.0	36.4	7.4	11.5	89.0	68.7	premium
15	100.0	44.4	15.4	43.9	26.1	7.6	7.8	84.0	98.0	regular
16	79.3	55.4	11.3	100.0	47.5	10.9	18.5	88.5	39.2	premium
17	100.0	56.2	16.9	39.6	18.7	6.3	6.5	77.0	57.7	regular
18	72.3	46.3	13.9	100.0	39.6	8.3	11.9	98.1	72.4	premium
19	100.0	49.5	16.0	49.4	38.0	10.5	13.1	82.8	65.7	regular
20	84.2	51.1	14.5	100.0	41.2	8.3	12.9	88.5	66.1	premium
21	100.0	48.5	15.6	50.6	30.4	9.2	10.8	84.7	79.4	regular
22	72.9	50.9	14.6	100.0	41.1	8.3	12.4	80.7	49.5	premium
23	100.0	56.5	16.5	40.6	20.1	7.3	6.8	79.4	57.0	regular
24	100.0	50.0	15.9	48.8	19.7	7.0	5.3	99.4	98.8	regular
25	86.3	54.4	12.9	100.0	45.0	11.2	17.9	79.4	40.0	premium
26	68.0	43.7	13.8	100.0	36.4	7.6	11.8	99.6	68.0	premium
27	100.0	43.5	13.6	59.3	37.6	9.3	13.1	72.2	50.1	regular
28	76.1	48.8	14.1	100.0	38.5	7.1	12.8	89.8	69.9	premium
29	100.0	49.8	16.6	48.4	30.5	8.8	11.7	85.0	76.3	regular
30	74.5	51.8	15.6	100.0	50.3	8.9	14.7	82.7	49.2	regular
31	100.0	59.8	17.8	31.3	17.5	6.7	5.2	82.5	64.4	regular
32	96.9	72.5	20.6	100.0	25.5	3.0	6.2	64.6	23.9	premium
33	100.0	51.0	13.8	62.7	27.3	5.5	8.5	88.4	77.3	regular
34	47.8	47.2	5.6	100.0	26.7	3.4	7.2	105.6	49.2	premium
35	100.0	47.9	16.8	56.5	35.0	11.1	12.7	89.1	84.1	regular
36	100.0	82.3	20.4	90.8	38.2	7.2	13.9	74.3	25.5	premium
37	100.0	58.0	19.3	43.9	24.4	9.9	8.2	81.8	60.0	regular
38	94.2	58.3	16.2	100.0	70.8	23.2	26.4	75.0	39.1	premium
39	100.0	55.0	17.9	41.0	18.7	6.2	5.4	86.7	72.4	regular
40	60.9	39.4	12.6	100.0	41.3	7.2	12.9	87.4	56.6	premium
41	100.0	54.4	18.2	54.3	36.7	10.7	12.0	85.3	68.4	regular
42	89.6	57.1	15.6	100.0	60.3	17.1	21.9	79.4	43.3	premium
43	100.0	46.7	15.0	58.9	39.0	8.1	13.1	77.4	44.5	regular
44	52.5	23.6	9.4	100.0	11.4	3.0	3.1	81.9	73.3	premium
45	100.0	59.6	20.5	65.5	28.5	10.2	10.9	76.2	52.7	regular
46	64.0	49.3	14.3	100.0	39.9	7.4	11.8	75.5	34.7	premium
47	100.0	52.3	17.4	60.9	39.0	8.9	14.1	83.7	60.2	regular
48	67.6	47.3	14.7	100.0	54.0	9.4	18.1	91.4	53.3	premium
49	100.0	52.1	17.1	37.9	43.6	9.0	16.0	85.1	69.3	regular
50	99.8	80.1	23.1	100.0	73.2	24.5	29.9	70.9	25.1	premium
51	73.2	62.4	16.4	100.0	84.1	24.9	31.0	72.1	22.0	premium
52	100.0	50.7	16.0	53.5	26.5	3.7	8.3	87.3	76.3	regular
53	69.8	52.9	15.0	100.0	34.3	3.9	9.3	77.4	24.0	premium
54	37.0	51.9	6.3	100.0	40.9	6.4	11.3	121.9	40.5	premium
55	100.0	48.0	16.4	49.7	26.9	4.0	7.9	86.0	79.6	regular
56	68.4	61.6	15.5	100.0	74.0	14.8	28.1	80.6	31.2	premium
57	59.9	48.5	14.6	100.0	54.4	10.0	16.9	74.3	33.4	premium
58	63.1	46.9	14.1	100.0	48.3	7.5	13.4	73.4	35.8	premium
59	40.2	27.5	14.0	100.0	66.3	15.2	24.9	78.6	56.0	premium
60	100.0	55.4	17.3	10.9	36.9	13.2	11.4	65.7	43.0	regular
61	100.0	53.6	14.7	53.1	14.5	2.6	3.0	72.5	45.7	regular

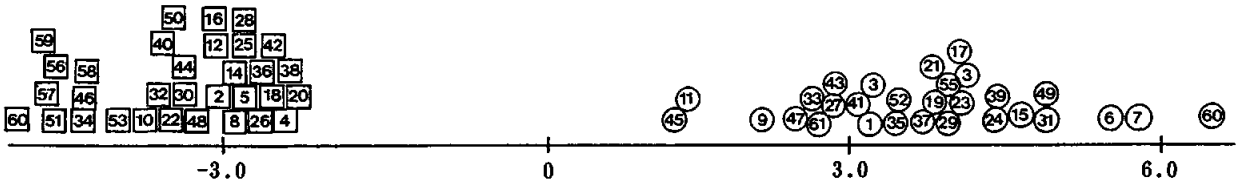

Fig. 2 Plotted canonical variates.

Table 4 *k*NN analysis of gasoline

	1NN	2NN	3NN	4NN	5NN
Total missed	0	0	0	0	0
Present correct	100	100	100	100	100

Table 5 Results of principal component analysis of MS data

Variable	Eigenvector 1	Eigenvector 2
M4	0.3472	0.5717
M1	−0.5668	−0.3574
M2	−4.8810	0.2742
M8	0.5125	−0.2180
M9	0.2393	−0.6501
Proportion	0.431	0.369

be used to distinguish between the two groups.

Application of the kNN method. The *k*NN method was applied to MS data of gasoline sample. The distance scale used here is the Euclidean distance. Table 4 shows the results, up to 5NN, of the evaluation of the prediction accuracy of this method from MS data of gasoline samples by the leave-one-out procedure.

Application of unsupervised learning pattern-recognition and related methods

Application of principal component analysis. Table 5 gives the results of applying a principal-component analysis to the MS data of gasoline. Since the cumulative proportion of the first and second principal components was 80.0%, the third principal component was disregarded. The principal component counts given to the samples are plotted on the first and second principal component axes in Fig. 3. Table 5 shows that the coefficient of the first principal component is for M4, M8 and M9 for positive values, and for M1 and M2 for negative values. It is thus seen that this axis represents the (aromatic and naphthenic compounds): (paraffin compounds) ratio. In Fig. 3, the premium gasolines are to the right of the regular gasolines. From this, the sign of this coefficient indicates that the greater is the content of aromatic and naphthenic compounds, the higher is the octane number. The coefficient of the second principal component is for M4 and M2 with a positive value and for M1, M8 and M9 with a negative value. Both M1 and M2 are caused by paraffinic compounds, the intensity ratio of which shows the structure difference of paraffinic compounds. It is thus considered that any structural differences of the paraffinic compounds are canceled from this axis due to the inverse sign, as well as the almost identical absolute values of M1 and M2. It is therefore found that the second principal component axis shows the (aromatic compounds): (naphthenic compounds and branched paraffins) ratio. It further seems to be the case that regular and premium gasolines have their own respective rules regarding mixtures of aromatic and naphthenic as well as branched paraffinic compounds. These apparent rules are indicated by the straight lines in the graph.

Application of MST. Figure 4 gives the results of a classification of gasolines by applying MST to the MS data. The greater length of the member between samples 11 and 20, 26 and 34, 24 and 35, 60 and 61 indicates that the demarcation among four classes can be positioned here. Again, the Euclidean distance is used as the distance scale.

MST, unlike cluster analysis, involves the fusion of samples at different intervals between them. By observing the combination modes of the longest edge in regular or premium gasoline (except for samples, 24, 34, 54 and 60) it can be seen that fusion takes place in the mode of the combination conforming to the mixture rule (represented by the straight line in Fig. 4) for aromatic, naphthenic and branched paraffinic compounds in both regular or premium gasolines, as stated above, with respect to the principal component analysis.

This study has revealed that motor gasolines can be distinguished from each another, or classified according to the octane number, by applying a pattern-recognition method to their MS spectra. The results of principal content analyses and MST have suggested that a rule exists that is based on the mixture ratio of aromatic and naphthenic and branched paraffin compounds in regular or premium gasolines.

Applying this method would presumably make it possible not only to determine whether an unidentified sample is regular or premium gasoline, but also what specific kind of gasoline it is among regular or premium gasolines, *i.e.* which of the gasolines sampled for this study it resembles.

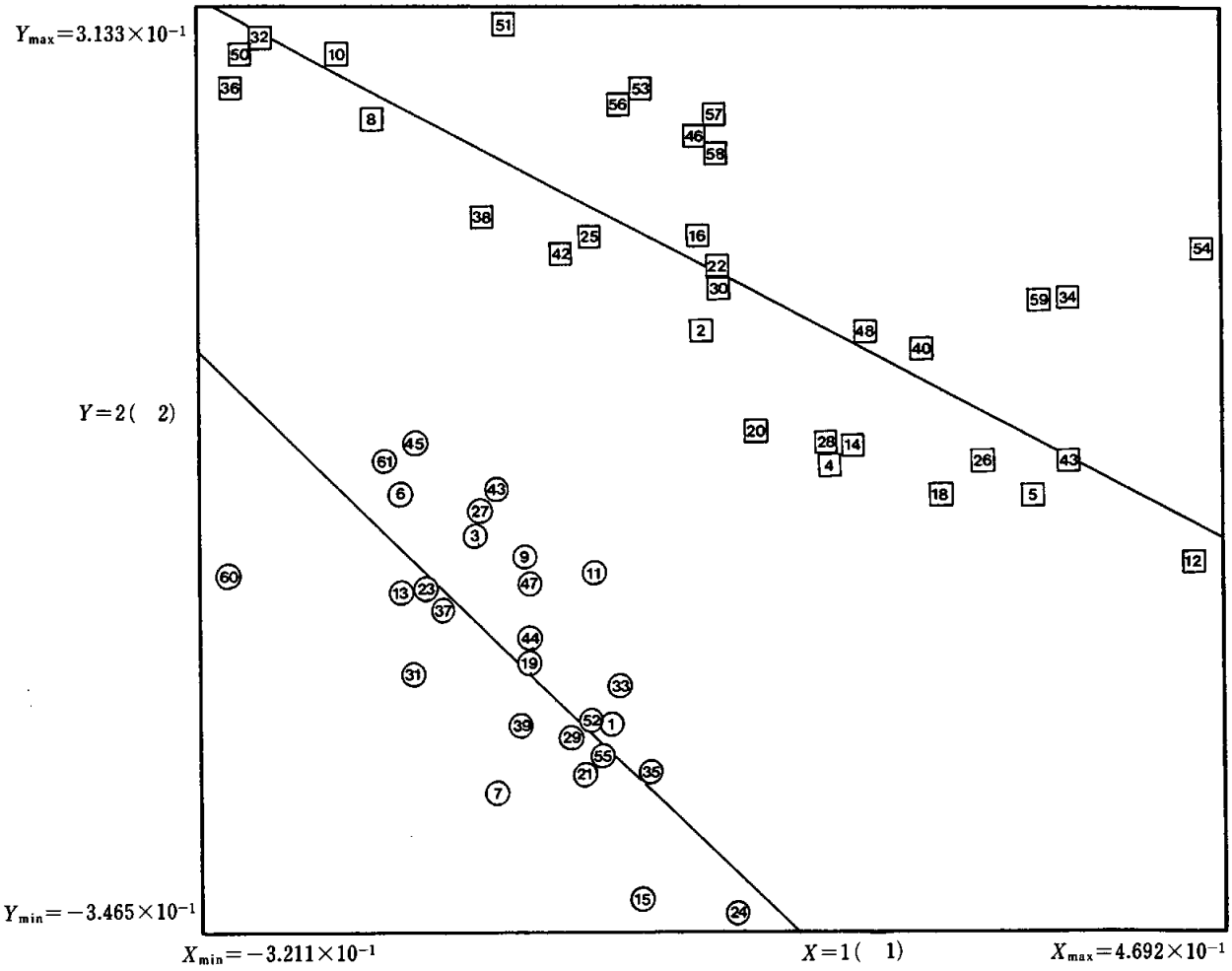

Fig. 3 Principal component counts of the plotted samples.

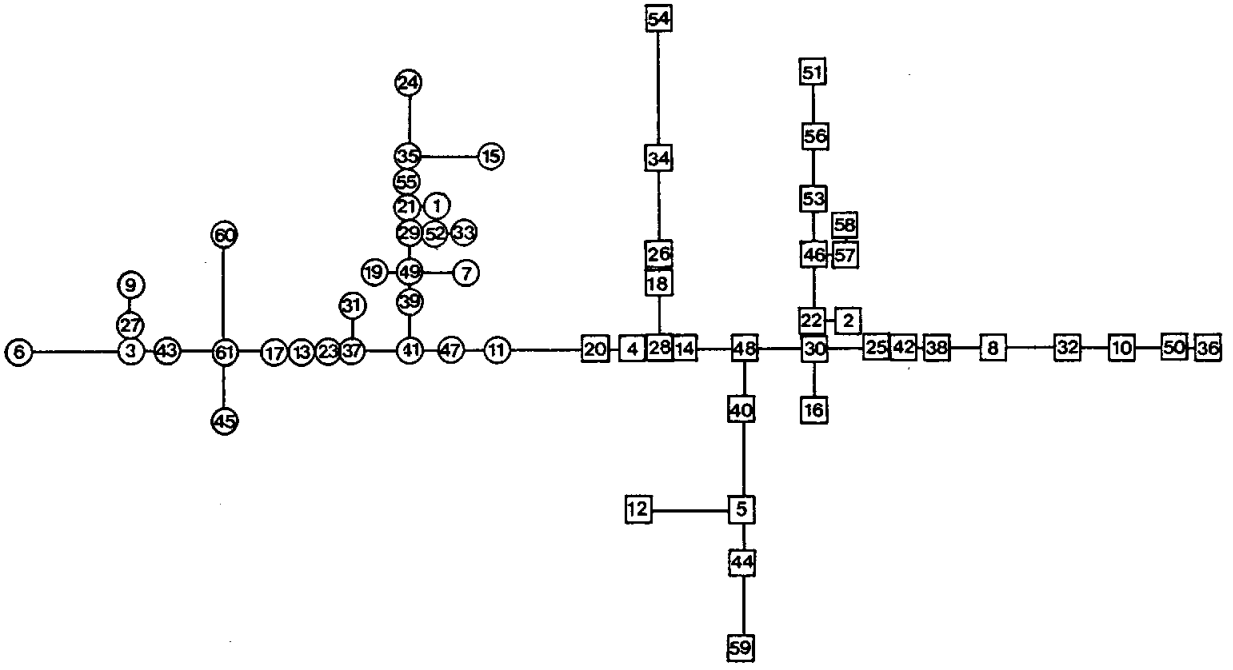

Fig. 4 MST results.

References

1. ASTM D2699, "*Standard Testing Method for Knock Characteristic of Motor Fuels by the Research Method*", American Society for Testing and Materials, Philadelphia, 1985.
2. M. E. Myers, J. J. Stollsteimer and A. M. Wims, *Anal. Chem.*, **47**, 2301 (1975).
3. J. Mühl and V. Srića, *Fuel*, **66**, 1146 (1987).
4. J. Mühl, V. Srića and M. Jednaćak, *Fuel*, **68**, 201 (1989).
5. P. C. Andarson, J. M. Sharkey and R. P. Walsh, *J. Inst. Petroleum*, 58 (560), **83** (1972).
6. A. Haas, G. McElhiney, W. Ginzel and A. Buchsbaum, *Sci. Technol.*, **1**, 21 (1990).
7. K. Ishihara and T. Yamashita, *Res. Col. Engin. Tokyo Univ.*, **25**, 103 (1989).
8. J. J. Kelly, C. H. Barlow, T. M. Jinguji and J. B. Callis, *Anal. Chem.*, **61**, 313 (1989).
9. M. Ichikawa, N. Nonaka, H. Amano, I. Takada and S. Ishimori, *Appl. Spectrosc.*, **45**, 637 (1991).
10. M. Ichikawa, N. Nonaka, I. Takada, S. Ishimori, H. Ando and K. Kumamoto, *Appl. Spectrosc.*, **45**, 1679 (1991).
11. M. Ichikawa, N. Nonaka, H. Amano, I. Takada, S. Ishimori, H. Ando and K. Kumamoto, *Appl. Spectrosc.*, **46**, 498 (1992).
12. M. Ichikawa, N. Nonaka, I. Takada and S. Ishimori, *Appl. Spectrosc.*, **46**, 966 (1992).
13. T. Okuno, T. Haga, H Kume and T. Yoshizawa, "*Tahenryoukaisekiho* (Multivariate Analysis, in Japanese)", p. 259, Nikkagiren, Tokyo, 1983.
14. H. Abe, *Kagakusousetu* Vol. 18, p. 119, Gakkaisyuppan Center, Tokyo, 1978.
15. K. Yakimoto, Y. Tanaka, "*Tahenryoukaisekiho* (Multivariate Analysis, in Japanese)", p. 31, Gendaisugakusha, Tokyo, 1982.
16. S. Sasaki, H. Abe, Y. Takahashi, C. Takayama and Y. Miyashita, "*Kagakusyanotameno pataanninshikijosetu* (An Introduction to Pattern Recognition for Chemists, in Japanese)", p. 65, Kagakudouzine, Tokyo, 1985.
17. "*ARTHUR User's Manual*", Chap. 5, p. 93, Infometrix Inc., Seattle, 1985.
18. "*BMDP User's Manual M Series, No. 8080-7-089-01*", Chap. 7, p. 145, Hitachi, Yokohama, 1978.
19. Y. Yoneda and H. Abe, "*Busseinoyosoku* (Prediction of properties, in Japanese)", Kagakutojoho (in Chemistry and Information), ed. M. Ohki, S. Sasaki and H. Chihara, Chap. 4, p. 91, Iwanami-syoten, Tokyo, 1981.
20. B. R. Kowalsky and C. F. Bender, *Pattern Recognition*, **8**, 1 (1976).
21. Y. Miyashita, M. Ishikawa and S. Sasaki, *J. Food Agric.*, **49**, 325 (1989).

(Received July 31, 1992)
(Accepted December 28, 1992)