COMMUNICATIONS

Determination of the Geographical Origins of Frozen Concentrated Orange Juice via Pattern Recognition

The geographical origins of frozen concentrated orange juice (FCOJ) were determined by application of the techniques of pattern recognition. Concentrations for 28 elements were determined in the FCOJ samples originating from Brazil and Florida. From these elements, five were chosen as target elements which were subsequently treated by pattern recognition. and proved to be more than adequate discriminators of geographical origin.

In the past several years, an increased interest in pattern recognition has arisen as is evidenced by the areas of its utilization. The earliest applications of pattern recognition to chemical data involved the determination of molecular structural features from low-resolution mass spectra. Subsequent applications were seen in such diverse fields as medicine, archaeology, aerial photography, fingerprint analysis, handwriting analysis, etc. In general, through great efforts pattern recognition (Jurs and Isenhour, 1975) has made untenable problems manageable. In this manuscript, the authors will present a novel application of pattern recognition, namely, to the elucidation of the geographical origin of frozen concentrated orange juice via trace elemental analysis.

For the past several years, McHard et al. (1976a,b) have been conducting a thorough investigation of the trace elemental composition of frozen concentrated orange juice (FCOJ) which had its origin in Florida, Brazil, Mexico, and California. The objective of their research efforts was to determine whether or not one can distinguish between a FCOJ produced in Florida and any or all FCOJ's produced outside of the state and if so to establish the methodology to do so. Recently, McHard et al. (1979) have submitted their results which clearly showed distinguishing features contained in the elements of Ba, Rb, B, Ga, and Mn. A subsequent statistical analysis of their results confirmed their ability to distinguish between the various geographical origins of FCOJ with a high level of confidence. An alternative to their statistical analysis is presented here which has the advantages of deleting the extensive statistical treatment of the data while retaining the decision making ability and its associated high level of confidence. Prior to giving a detailed description of this work, a brief review of the basics of pattern recognition is given. For a more detailed treatment of the theory, the reader is referred to the works of Kowalski (1975) and Jurs and Isenhour (1975).

THEORY

A simple physical interpretation of the essence of pattern recognition will serve as a convenient starting point for our discussion. Essentially, we are interested in describing geometric regions in space and then finding a means to separate them. The regions in space, or categories, are described by "features" of the measurement process. For instance, if five elements are to be measured, (A, B, C, D, and E), we are dealing with a five-dimensional space whose magnitude along any axis is governed by the measured concentration for that element. We now can define a pattern vector for one sample set (one particular FCOJ sample):

$$\mathbf{X} = (A, B, C, D, E) \tag{1}$$

This vector now represents a single point in five-dimensional space that completely characterizes that one particular FCOJ sample. If we now make the assumption that similar FCOJ samples can be described by pattern vectors that define points in a similar region in space, we now have a basis upon which to form categories. The problem reduces to one in which we are interested in discriminating between these categories, or more simply we are looking for a means of classifying sample sets into one of several categories. The means by which this is accomplished is by the introduction of a "decision vector". The decision vector (eq 2) is an empirically derived vector (which can be improved with appropriate training schemes; Kowalski, 1975) which upon multiplication with the sample vector (eq 3) will yield one of two results

$$\mathbf{D} = (a, b, c, d, e) \tag{2}$$

$$\mathbf{X}\cdot\mathbf{D} = (Aa + Bb + Cc + Dd + Ee) = S \tag{3}$$

where $S > 0$ or $S < 0$. If $S > 0$, then X belongs to category I, if $S < 0$, then X belongs to category II. Physically, the decision vector describes a surface that divides the two categories in question. By multiplication of a sample vector by the decision vector, one is really making a decision as to which side of the decision surface the sample vector lies.

The utility of pattern recognition resides within the ability of the decision vector to discriminate between two categories. Having no a priori knowledge of the sample set imposes great restrictions on the technique. However, utilization of whatever information is available can greatly reduce the number of iterative steps needed in order to arrive at a satisfactory decision vector. In the end, the efficiency of a decision vector is determined by repeated attempts at making the discriminating decision and recording the success rate.

RESULTS AND DISCUSSION

By utilization of the results in Table I, ,the proposed pattern recognition method was applied in discriminating Floridian FCOJ from Brazilian FCOJ. The five target elements chosen were identical with those chosen by McHard et al. (1979), i.e., Ba, B, Ga, Mn, and Rb. This was done to allow a direct comparison between the statistical decision method used in the previous analysis and the newly proposed method presented here. By careful

0021-8561/80/1428-1306$01.00/0

Table I.[a] Concentration Ranges for the Five Target Elements Analyzed in Brazilian and Floridian Samples of FCOJ[a]

	Florida concn range, ppm	Brazil concn range, ppm
Ba	0.025–0.07 ($\bar{x}$ = 0.048)	0.139–0.53 ($\bar{x}$ = 0.33)
B	0.95–1.20 ($\bar{x}$ = 1.08)	0.675–1.61 ($\bar{x}$ = 1.14)
Ga	0.03–0.04 ($\bar{x}$ = 0.035)	0.047–0.11 ($\bar{x}$ = 0.079)
Mn	0.25–0.315 ($\bar{x}$ = 0.28)	0.30–0.60 ($\bar{x}$ = 0.45)
Rb	0.365–0.740 ($\bar{x}$ = 0.55)	1.98–4.86 ($\bar{x}$ = 3.42)

[a] All values reported are adjusted for variations in solid content and express concentrations found in single strength orange juice; data taken from McHard et al. (1979).

inspection of the data and a minimal amount of trial and error, the following decision vector was arrived at:

$$\mathbf{D} = (1, -1, -1, -1, 1) \quad (4)$$

The decision vector was formulated such that when the dot product of it and a sample vector is computed, if

$S < 0$, then X is a Florida juice

$S > 0$, then X is a Brazilian (or Mexican) juice

In general, trace elemental concentrations in juices of Floridian origin were lower than those of nondomestic origin. Therefore, the case in which we are forced to distinguish between a Florida juice with its highest possible values for the target elements and a Brazil juice with its lowest values would clearly demonstrate the effectiveness of the chosen decision vector for the worst possible case. It should be mentioned, at this point, that all values for elemental concentrations should be referred back to single strength orange juice from the FCOJ. A correction for solid content variations must be included as well. However, as pointed out by McHard et al. (1979), a convenient method for bypassing the need to determine the solid content separately for each sample was to make use of elemental ratios. Zinc was chosen as the reference element to which all other elements were ratioed. For details on the ratioing procedure, the reader is referred to the original manuscript (McHard et al., 1979). The actual values used in the following two cases can be found in Table I.

Case I: *Case Ia.* Here, $\mathbf{X}_1$ = (0.07, 1.20, 0.04, 0.315, 0.740) (Floridian), where the vector represents concentrations (in ppm) for the following elements in the designated appropriate order, namely, Ba, B, Ga, Mn, and Rb, and the decision vector is given by eq 4. Now we must evaluate $S = \mathbf{X}\cdot\mathbf{D}$, and so

$$\mathbf{X}_1\cdot\mathbf{D} = (0.025 - 0.95 - 0.03 - 0.25 + 0.37) = S$$

$$S = -0.84$$

Since $S < 0$, therefore $\mathbf{X}_1$ is a Florida juice.

Case Ib. Here, $\mathbf{X}_2$ = (0.53, 1.61, 0.11, 0.60, 4.86) (Brazilian). Proceeding as before, $S_2 = \mathbf{X}_2\cdot\mathbf{D}$ = 307. In this case, $S_2 > 0$ and therefore $\mathbf{X}_2$ is a Brazilian juice.

Case II. To parallel the presentation of McHard et al. (1979), case I will be repeated by using elemental ratios which were ratioed to zinc and can be found in the original reference (McHard et al., 1979). This required a modification of our decision vector. However, the decision rules still remain the same. The new decision vector is given by

$$\mathbf{D}_n = (1, -2, -2, -2, 1) \quad (5)$$

and so cases IIa and IIb are as follows.

Case IIa. $\mathbf{X}_1$ = (2.9661, 0.8702, 1.2812, 1.0586, 1.8473) (Florida)

$$\mathbf{D}_n\cdot\mathbf{X}_1 = -1.603$$

$$S_1 < 0 \therefore \mathbf{X}_1 \text{ is a Florida juice}$$

Case IIb. $\mathbf{X}_2$ = (5.150, 0.9448, 1.3794, 1.3816, 6.0511) (Brazil)

$$\mathbf{D}_n\cdot\mathbf{X}_2 = 3.789$$

$$S > 0 \therefore \mathrm{X}_2 \text{ is a Brazil juice}$$

CONCLUSIONS

It has been shown that for the worst possible case in which *all* elements of interest in a Florida FCOJ have their highest possible values and those of the Brazil FCOJ have their lowest values, the proposed technique offers a clear and substantial separation of the two groups. This in actuality represents the worst possible case to be expected in that very rarely will we encounter a sample that possess concentrations of all the target elements stacked to one side of the mean. We therefore feel very confident that the derived decision vectors will allow adequate discrimination between various geographical locations. Assignment of a success rate of our decision vector would entail performing x number of trials with it and monitoring the number of correct reponses. With the sample of the population we have obtained to date, our decision vector is 100% successful. The advantage offered by this method of data analysis and decision making lies in the reduced number of mathematical operations performed as well as the omission of any direct utilization of statistical decision theory.

LITERATURE CITED

Jurs, P. C., Isenhour, T. L., "Chemical Applications of Pattern Recognition", Wiley, New York, 1975.
Kowalski, B. R., *Anal. Chem.* **47**, 1152A (1975).
McHard, J. A., Winefordner, J. D., Attaway, J. A., *J. Agric. Food Chem.* **24**, 41 (1976a).
McHard, J. A., Winefordner, J. D., Ting, S. U., *J. Agric. Food Chem.* **24**, 950 (1976b).
McHard, J. A., Winefordner, J. D., Foulk, S. F., *J. Agric. Food Chem.* **27**, 1326 (1979).

S. Bayer
J. A. McHard
J. D. Winefordner*

Department of Chemistry
University of Florida
Gainesville, Florida 32611

Received for review November 19, 1979. Accepted May 2, 1980. This work was supported by Air Force Grant F44620-76-C-0005.