A NEW APPROACH FOR ORDERED MULTICATEGORICAL CLASSIFICATION USING SIMPLEX TECHNIQUE

Yoshimasa TAKAHASHI, Yoshikatsu MIYASHITA, Hidetsugu ABE,
and Shin-ichi SASAKI[®]

Department of Materials Science, Toyohashi University of
Technology, 1-1, Hibarigaoka, Tempaku, Toyohashi 440

A new approach to the ordered multicategorical classification has been developed. The approach is based on the pattern recognition technique using simplex optimization. The ORMUCS method gives a single discriminant function for classification of the ordered multicategorical data. The ORMUCS method was compared with adaptive least square method (ALS method) in application to the classification problem of the ordered potency classes of antitumor mitomycin derivatives. In the full analysis and leave-one-out test, the ORMUCS method proved itself to be more stable than the ALS method for developing the discriminant function.

INTRODUCTION

In recent years, quantitative procedures have been applied to the structure-activity studies of biologically active compounds. The Hansch approach¹⁾ is most widely used for the study, especially, for the estimation of potency of biological activity with continuous scale. The potency is often represented in either ordinal scale or a form of activity rating as well as in the form of continuous scale. To such a discrete type of biological activity data, the Hansch approach is not much suitable. Thus, values in the ordinal scale should not be considered as the value for the intrinsic activity of each compound, but as an ordered category for the level of potency. Considering the matter, classification techniques in pattern recognition may be applicable to solve such problem.

Moriguchi et al. presented a technique (ALS) to be applied to structure-activity correlation studies using such ordered categorical data. But, there are several hard problems, in ALS, on the procedures of developing the best discriminant function for class discrimination. (1) The final solution mostly depends on the number of iterations because of the nonconvergency of algorithm of the ALS method. (2) The approach is based on the iterative correction of activity score whereas the meanings of constant parameters used in the correction equation²⁾³⁾ are not well defined and the values to be strictly specified are accompanied with considerable arbitrariness.

In the present study, considering the above problems, the authors have developed a new approach, ORMUCS (Ordered Multicategorical Classification using Simplex

optimization techniques) method, for structure-activity correlation studies with ordered multicategorical data. The applicability of the ORMUCS method will be described in comparison with the ALS method.

EXPERIMENTAL AND RESULTS

ORMUCS method

The procedure of ORMUCS method is outlined in Fig. 1. In the method, the discriminant function for classification is described by equation 1.

$$S(\mathbf{X}) = \sum_{i=1}^d w_i x_i = \mathbf{W}\mathbf{X} \quad (1)$$

where x_i is the i -th component of $\mathbf{X}$ which is represented as a pattern vector in the d -dimensional measurement space, and w_i is the weight assigned to the component. The vector described by set of the weights ($w_1, w_2, \dots, w_d$) is called as weight vector. $S(\mathbf{X})$ is the discriminant score to be used for classification of the pattern vector $\mathbf{X}$.

Fig.1 Ordered Multicategorical Classification using Simplex technique (ORMUCS)

The rules of the classification in ORMUCS are defined by the equations below.

$$C_{\min.}^{(k)} \leq S(\mathbf{X}) < C_{\max.}^{(k)} \quad \text{implies } \mathbf{X} \in \text{class } k \quad (2)$$

where $C_{\min.}^{(k)}$ and $C_{\max.}^{(k)}$ are lower and upper limits of the discriminant score for category k , respectively. These rules are also adopted to the development of the discriminant function.

In the present method, the simplex optimization technique is employed in developing the discriminant function. Firstly, an initial weight vector should be given by an ordinary least squares method.

$$\begin{array}{rcl}
 A_1 & = & b_1x_{11} + b_2x_{12} + \cdots + b_dx_{1d} \\
 \vdots & & \vdots \\
 A_i & = & b_1x_{i1} + b_2x_{i2} + \cdots + b_dx_{id} \\
 \vdots & & \vdots \\
 A_n & = & b_1x_{n1} + b_2x_{n2} + \cdots + b_dx_{nd}
 \end{array} \quad (3)$$

In equation 3, A_i is the activity score of the i -th compound. In this method, for convenience sake, the middle value of the range of class i which was defined in equation 2 is used for the activity score. The pattern vector corresponding to the i -th compound consists of the components, $x_{i1}, x_{i2}, \dots, x_{id}$.

The least squares estimates of b_k ($k=1,2,\dots,d$) computed from equation 3 are used as the components of a initial weight vector, as shown below.

$$W_1 = (w_1, w_2, \dots, w_d) = (b_1, b_2, \dots, b_d) \quad (4)$$

In the next step, the other d weight vectors must be chosen for making the simplex to be spanned in the d -dimensional weight space. Generation of them is as follows,

$$\begin{array}{rcl}
 W_1 & = & (w_1, w_2, \dots, w_{d-1}, w_d) \\
 W_2 & = & (w_1 + \delta_1, w_2, \dots, w_{d-1}, w_d) \\
 W_3 & = & (w_1, w_2 + \delta_2, \dots, w_{d-1}, w_d) \\
 \vdots & & \vdots \\
 W_d & = & (w_1, w_2, \dots, w_{d-1} + \delta_{d-1}, w_d) \\
 W_{d+1} & = & (w_1, w_2, \dots, w_{d-1}, w_d + \delta_d)
 \end{array} \quad (5)$$

where δ_j was defined as

$$\delta_j = \alpha w_j \quad (j = 1, 2, \dots, d) \quad (6)$$

In equation 6, α is a constant parameter related with the size of the initial simplex. In the present study, this value is defined to be between 0.1 and 1.0.

Besides, a prototype weight vector (initial weight vector) W_1 , in generating the initial simplex, was determined by the use of the least squares method.

Once the initial simplex was placed in the objective space, optimization of the simplex (or weight vectors) may be attempted by a certain moving algorithm (Fig. 2). The simplex in the optimization process is moved along the algorithm suggested by Nelder and Mead⁴⁾, which is a logical algorithm consisting of reflection,

Fig. 2 Progress of 2-dimensional simplex toward optimum area

expansion and contraction rules. This procedure has been already described in detail⁵⁾⁻⁷⁾. Here, it will be still necessary to give some explanation, using Fig. 3. Because the procedure makes a major part of ORMUCS.

In Fig. 3, suppose W in the initial simplex BMW gives the worst response, B gives the best response and M an intermediate response. After evaluation of the response at each point the point, W, which gives the least desirable response, is discarded and another new point is placed. Then, a new weight vector corresponding to the new point is given as follows

$$\mathbf{W}_{\text{new}} = \mathbf{W}_P + \gamma(\mathbf{W}_P - \mathbf{W}_W) \quad (7)$$

where $\mathbf{W}_P$ is the centroid vector defined by the remaining points but for W. Reflection: In Fig. 3, P is the centroid of the remaining points excepting for W. Reflection of W through the centroid point generates a new point R. The new weight vector $\mathbf{W}_R$ corresponding to the point R is given by equation 7 with $\gamma=1$. Expansion: If the response at R is more desirable than that at B, another new point, E is placed by equation 7 with $\gamma=2$. This implementation is called expansion. Then, if the response at E is better than those at B and M, the point E is kept to form a new simplex using BME. Otherwise, the new point is discarded and a point R is kept to form the new simplex. Contraction: If the response at R is less than that at M, contraction is employed in the following two cases. a) If the response at R is better than at W, at least, the contraction is implemented in the direction nearby R. The new point C is given by equation 7 with $\gamma=\frac{1}{2}$ followed by the formation of new simplex by BMC. b) If the response at R is less desirable than at W a new point is placed nearby W. The new point, T, is given by equation 7 with $\gamma=-\frac{1}{2}$. Then, the new simplex is formed by BMT.

For another cases, normal reflection is employed in the simplex moves. In that case, thus, the new simplex is formed by BMR. These procedures in the simplex moves are repeated until any final convergence criterion is satisfied.

Criteria for the best weight vector

Two criteria were employed for the evaluation of each weight vector corresponding to vertex on the simplex. One is recognition rate, R, as shown below.

$$R = N_c/N_t \quad (8)$$

where, N_c is the number of samples correctly classified using the weight vector, and N_t is the total number of the samples. The other is perceptron function, J. Definition of J is as follows.

$$J = \sum_{\mathbf{x}_i \in S_E} |\mathbf{W} \cdot \mathbf{x}_i^{(k)} - C_{\text{lim}}^{(k)}| \quad (9)$$

Fig. 3 Rules for the movement of simplex.

BMW is a initial simplex, R is the reflection point, E is the expansion point, C and T are the positive and the negative contraction point respectively.

where

$$C_{\text{lim.}}^{(k)} = \begin{cases} C_{\text{min.}}^{(k)} & \text{for } \mathbf{W} \cdot \mathbf{X}_i^{(k)} < C_{\text{min.}}^{(k)} \\ C_{\text{max.}}^{(k)} & \text{for } \mathbf{W} \cdot \mathbf{X}_i^{(k)} \geq C_{\text{max.}}^{(k)} \end{cases} \quad (10)$$

In equation 9, $\mathbf{W}$ is the weight vector to be evaluated and $\mathbf{X}_i$ is an element of the set of patterns which were misclassified, S_E . The k is the class to which the pattern $\mathbf{X}_i$ belongs and, $C_{\text{lim.}}^{(k)}$ is the limited value of the class k .

Among the weight vectors, the value of R is the larger the more desirable. If the values of R for two or more weight vectors are equal to one another, the one with the smallest value of J should be selected as the best weight vector. In general, the value of J will become smaller in accordance with the increase of misclassification of the compounds.

First criterion, ie, recognition rate, gives the discontinuous response surface. Therefore the ultimate solution will be given by the minimum J value for weight vectors giving maximum recognition rate. In ORMUCS, these criterions are used as the final convergence conditions in the simplex optimization process. The optimization process ceases when the value of R becomes to 1.0 (or $j=0$), or when the R and J criterions reached a steady state. We call the state that the change of J value among ten consecutive moves with equal recognition is less than 0.1% with respect to the first J value in that implementation "a steady state".

Application to the study of a structure-activity problem

To compare the ability of ORMUCS with ALS²⁾, our method has been applied to the study of structure-antitumour activity correlations of mitomycin derivatives. Table 1 shows the data set used in this implementation. Five descriptors were used in order to describe each compound. They are field constant, F_x , and Van der Waals volume, V_{w-x} , for substituent X , Taft's steric substituent constant, E_{s-z} , for substituent Z , and two dummy parameters Y_{OMe} and Y_{OH} , accounting for substituent $Y = \text{OMe}$ and $Y = \text{OH}$, respectively.

Mitomycin derivatives were categorized into five groups in accordance with the potency of antitumour activity of each of them. The boundaries of activity classes were assigned to each class as follows,

$$\begin{array}{lll} \text{if} & S(\mathbf{X}) < 1.0 & \text{then } \mathbf{X} \in \text{class 1 } (-) \\ \text{if} & 1.0 \leq S(\mathbf{X}) < 2.0 & \text{then } \mathbf{X} \in \text{class 2 } (+) \\ \text{if} & 2.0 \leq S(\mathbf{X}) < 3.0 & \text{then } \mathbf{X} \in \text{class 3 } (2+) \\ \text{if} & 3.0 \leq S(\mathbf{X}) < 4.0 & \text{then } \mathbf{X} \in \text{class 4 } (3+) \\ \text{if} & 4.0 \leq S(\mathbf{X}) & \text{then } \mathbf{X} \in \text{class 5 } (4+) \end{array} \quad (11)$$

The conditions in this study are all the same as these used in ALS²⁾.

The best weight vector and recognition rate which were obtained by ORMUCS are shown comparing with ALS results. As shown in Table 2, the recognition rate, 93.8 %, obtained by ORMUCS method is equal to the ALS results. Only one compound belonging to class 1 (compound 15 in Table 1) was misclassified as class 3. The detail of the misclassification is also the same as the result of ALS analysis. Then,

Table 1 Mitomycin derivatives and descriptors used for the analysis^{a)}

No.	X	Y	Z	F _x	V _{w-x}	Y _{OMe}	Y _{OH}	E _{s-z}	Act.
1	NH	OMe	H	0.02	0.177	1	0	1.24	3+
2	NH ₂ t	OMe	H	-0.11	0.493	1	0	1.24	3+
3	NH	OMe	Me	0.02	0.177	1	0	0	2+
4	NH	OMe	Et	0.02	0.177	1	0	-0.07	2+
5	NH	OMe	Ac	0.02	0.177	1	0	-0.47	2+
6	NH	OH	Me	0.02	0.177	0	1	0	2+
7	NMe	OMe	H	0.10	0.441	1	0	1.24	2+
8	NH	OMe	COPh-o-Cl	0.02	0.177	1	0	-1.19	+
9	NH	OMe	COPh-p-Cl	0.02	0.177	1	0	-1.19	+
10	NHPh	OMe	H	-0.02	0.892	1	0	1.24	+
11	OMe	OMe	H	0.26	0.304	1	0	1.24	+
12	OMe	OMe	Me	0.26	0.304	1	0	0	+
13	OMe	OH	Me	0.26	0.304	0	1	0	±
14	NH	H	Me	0.02	0.177	0	0	0	-
15	NH	OMe	SO Me	0.02	0.177	1	0	-1.54	-
16	OMe	H	Me	0.26	0.304	0	0	0	-

a) Taken from ref.2.

Table 2 Comparison of the discriminant functions between ORMUCS and ALS with the antitumor activity data of 16 mitomycins

Method	F _x	V _{w-x}	Y _{OMe}	Y _{OH}	E _{s-z}	Const.	Recog. ^{b)} (%)	Rs ^{c)}
ORMUCS	-4.40	-2.54	2.66	2.51	0.81	1.26	93.8	0.963
ALS ^{a)}	-4.33	-2.46	2.48	2.28	0.77	1.42	93.8	0.963

a) Taken from ref.2. b) Recog; recognition rate.

c) Rs; Spearman's rank correlation coefficient.

it may be considered that the differences in the components of the two weight vectors in Table 2 suggest the flexibility of the solution in the neighbourhood of the optimum area. The recognition resulted by the use of ORMUCS is shown in Table 3, with the value of discriminant function.

Consideration by leave-one-out predictions

In the predictive test using leave-one-out technique, a comparison of ORMUCS method with ALS has been done. Leave-one-out predictions by ALS method were not reported in the reference(2). Therefore, the leave-one-out prediction test with ALS was implemented by the use of the program which was developed by the authors themselves according to the algorithm in references(2,3). Validity of the program developed was justified with the consistency to the results²⁾³⁾ reported by Moriguchi using ALS program in full analysis.

The prediction results by both the methods are shown in Table 4. As the results, three compounds were misclassified by ORMUCS method, whereas, by ALS method, eight compounds were misclassified. Considering the ordered classes of the

Table 3 Recognition results by ORMUCS^(a)

No.	Class	Func. Value	Recog.
1	5	4.378	5
2	5	4.149	5
3	4	3.379	4
4	4	3.322	4
5	4	3.000	4
6	4	3.231	4
7	4	3.356	4
8	3	2.419	3
9	3	2.419	3
10	3	2.740	3
11	3	2.999	3
12	3	2.000	3
13	2	1.852	2
14	1	0.717	1
15	1	2.137	3 (b)
16	1	-0.661	1

(a) Using the weight vector in Table 2.

(b) misclassification

Table 4 Comparison between ORMUCS and ALS with the antitumor activity data of 16 mitomycins by the leave-one-out predictions

Compd.	Class	ORMUCS	ALS ^{a)} (30)	ALS ^{a)} (100)
1	(5)	5	5	5
2	(5)	5	5	5
3	(4)	4	4	4
4	(4)	4	4	4
5	(4)	4	3 *	3 *
6	(4)	4	3 *	3 *
7	(4)	4	4	4
8	(3)	3	3	3
9	(3)	3	3	3
10	(3)	3	4 *	4 *
11	(3)	4 * ^{b)}	5 *	5 *
12	(3)	2 *	2 *	2 *
13	(2)	2	3 *	3 *
14	(1)	1	2 *	2 *
15	(1)	3 *	3 *	3 *
16	(1)	1	1	1

a) The figure in parenthesis shows the number of iterations.

b) Asterisk shows a misclassified sample.

compounds misclassified, it will be observed in Table 4 that the two compounds excepting compound 15 were misclassified into neighbour class by ORMUCS, while by ALS method compound 11 which was a member of class 3 was also misclassified into a far apart class, ie, class 5. Then as shown in Table 3, the compounds excepting compound 15 misclassified by ORMUCS are critical samples. Compound 11 is a upper critical sample in class 3 and compound 12 in a lower critical one in that class. The fact shows that the misclassifications presented by leave-one-out test with ORMUCS might been possibly anticipated. Consequently, these results suggest that ORMUCS method is more stable than ALS method from the viewpoint of developing the best discriminant function in the case above-mentioned.

CONCLUSION

The ORMUCS method for ordered multicategorical classifications has been developed. Specifications of the present method are summarized as follows, 1) a single discriminant function is given, 2) this method is one of the distribution free approaches, 3) the weight vector of discriminant function is directly searched for based on the simplex optimization technique, and 4) the present method is more stable than the ALS method for developing the discriminant function.

The authors are grateful to the computer center, Institute for Molecular Science for affording facilities for computation.

REFERENCES

- 1) C.Hansch, T.Fujita: J.Am.Chem.Soc., 86, 1616(1964).
- 2) I.Moriguchi, K.Komatsu: Chem.Pharm.Bull., 25, 2800(1977).

- 3) I.Moriguchi, K.Komatsu, Y.Matsushita: J.Med.Chem., 23, 20 (1980).
- 4) J.A.Nelder, R.Mead: Comput.J., 7, 308 (1965).
- 5) S.N.Deming, S.L.Morgan: Anal.Chem., 45, 278a (1973).
- 6) L.Ritter, S.R.Lowry, C.L.Wilkins, T.L.Isenhour: Anal.Chem., 47, 1951 (1975).
- 7) Y.Takahashi, Y.Miyashita, Y.Tanaka, H.Hayasaka, H.Abe, S.Sasaki: J.Pharm.Sci.,
in press.

Keyword phrases

ordered multicategorical classification; simplex pattern recognition; structure-activity studies of biological active compounds.

(Received July 12, 1984)