Nonlinear Discriminant Analysis of Categorical Data by Fuzzy c-Means Fixed Point Iteration *

Hidetomo Ichihashi, Ryuichi Niiyama, Katsuhiro Honda, Chi-Hyon Oh
Graduate School of Engineering, Osaka Prefecture University
1-1 Gakuencho, Sakai, Osaka 599-8531 Japan
ichi@ie.osakafu-u.ac.jp

Abstract

In this paper, we introduce memberships in fuzzy clusters to the discriminant analysis for categorical data, which is known as Hayashi's second quantification method. Cluster partition is unknown and is determined by the combined algorithm of the Hayashi's discriminant analysis and the Fuzzy c-Means(FCM) clustering. The multiplication of unknown memberships makes the underlying models nonlinear ones. Contrary to FCM, the clustering criterion is not the distance from cluster center but is the between group variation. The fuzzy clusters are implicitly used for classification of new data.

Keywords: Fuzzy clustering, categorical data, quantifying method, nonlinear discriminant analysis

1 Introduction

Nominal measurement consists of assigning objects to categories. Variables measured on a nominal scale are often referred to as categorical or qualitative variables. The discriminant analysis for categorical data is known as Hayashi's second quantification method [1]. Individuals usually respond to a questionnaire by selecting category in each item of questions. Individuals are divided into some groups or classes by their respective external criteria or characters. The objective of the analysis is to predict a group to which a newly given individual belongs. Like other data analysis methods, many of the quantification methods for categorical data are developed on a linear model. In this paper we introduce memberships in fuzzy clusters to the discriminant analysis for categorical data. We use the term "cluster" separately from class or group. Cluster partition is unknown to all individuals and is determined by the

*Proc. of Joint 1st International Conference on Soft Computing and Intelligent Systems and 3rd International Symposium on Advanced Intelligent Systems, #24B4-1 (2002)

Table 1: Data format for discriminant of categorical responses

group number	individual	item 1	...	item R
		1 ... x_1	...	1 ... x_R
1	1	√		√
	⋮	√	⋮	√
	n_2			
⋮	⋮			
K	1	√		√
	⋮	⋮	⋮	⋮
	n_k	√		√

combined algorithm of discriminant and cluster analysis. The multiplication of memberships (i.e., unknown parameters) makes the underlying models nonlinear ones. Fuzzy c-Means (FCM) clustering by Bezdek *et al.* [2] is the popular method that partitions a data set into fuzzy clusters. Though we adopt this approach for clustering, the clustering criterion is not the distance from cluster center but the between group variation. In our approach, the groups or classes are crisp but the clusters are fuzzy, and the fuzzy clusters are implicitly used for classification of new individuals. Numerical example shows improvement in classification performance.

2 Simultaneous Application of Clustering and Quantification Method

Table 1 shows the example of questionnaire for discriminant analysis of categorical data. Responses of each individual are listed and individuals are divided into K classes. To represent the responses we introduce following dummy variables.

$$\delta_i p(jk) = \begin{cases} 1 & \text{individual p responded} \\ & \text{to category k of item j} \\ 0 & \text{else} \end{cases} \tag{1}$$

Each individual is given a value Y_{cip} by a weighted linear sum of the dummy variables as:

$$Y_{cip} = \sum_{j=1}^{R} \sum_{k=1}^{x_j} a_{cjk} \delta_{ip}(jk) \tag{2}$$

In Eq.(2), coefficients a_{cjk} denote the weight for category k of item j in cluster c. It should be noted that the individuals are divided into K groups or

classes. The group of each individual is assumed to be known and its membership to the group is crisp. The individuals are also partitioned into C clusters that are unknown and the memberships are fuzzy. The membership values are assigned by the fixed point iteration algorithm in FCM. To clarify the differences of meaning between the "class" and "cluster", we use the term "group" instead of "class". The objective function is formulated so as to discriminate all individuals into K groups by the values of Eq.(2). For this purpose Hayashi's second quantification method maximizes the between-group variation against the within-group variation and the weight coefficients a_{cjk} are determined to attain this goal. We follow this approach and introduce membership u_{cip} to change the linear model Eq.(2) into nonlinear one. The subscripts c, i and p represent the cluster number, group number and individual number respectively.

Though we introduce the membership u_{cip}, the variation of Y_{cip} can be decomposed as is known in the analysis of variance (ANOVA).

$$\begin{aligned}\sum_{c=1}^{C}\sum_{i=1}^{K}\sum_{p=1}^{n_i} u_{cip}(Y_{cip}-\bar{Y}_c)^2 \\ = \sum_{c=1}^{C}\sum_{i=1}^{K} n_{ci}(\bar{Y}_{ci}-\bar{Y}_c)^2 \\ + \sum_{c=1}^{C}\sum_{i=1}^{K}\sum_{p=1}^{n_i} u_{cip}(Y_{cip}-\bar{Y}_{ci})^2 \end{aligned} \tag{3}$$

The first and the second terms of the right hand side represent the between-group variation S_{B_c} and the within-group variation S_{W_c}. u_{cip} is constrained by

$$\sum_{c=1}^{C} u_{cip} = 1, \quad u_{cip} \in [0,1], \quad p = 1, \cdots, n_i, i = 1, \cdots, k \tag{4}$$

C denotes the total number of clusters. n_{ci} is the sum of the memberships of individuals in group i.

$$n_{ci} = \sum_{p=1}^{n_i} u_{cip} \tag{5}$$

$\bar{Y}_c$ is the grand mean of Y_{cip} in cluster c.

$$\bar{Y}_c = \frac{\sum_{i=1}^{K}\sum_{p=1}^{n_i} u_{cip}Y_{cip}}{\sum_{i=1}^{K}\sum_{p=1}^{n_i} u_{cip}} = \frac{\sum_{j=1}^{R}\sum_{k=1}^{x_j} n_{cjk}a_{cjk}}{n_c} \tag{6}$$

n_{cjk} is the sum of the memberships of individuals who responded to the category k of the item j.

$$n_{cjk} = \sum_{i=1}^{K} \sum_{p=1}^{n_i} u_{cip} \delta_{ip}(jk) \tag{7}$$

$\bar{Y}_{ci}$ is the within group mean in each cluster.

$$\bar{Y}_{ci} = \frac{\sum_{p=1}^{n_i} u_{cip} Y_{cip}}{\sum_{p=1}^{n_i} u_{cip}} = \frac{\sum_{j=1}^{R} \sum_{k=1}^{x_j} g_c^i(jk) a_{cjk}}{n_{ci}} \tag{8}$$

$g_c^i(jk)$ is the sum of the memberships in cluster c of individuals in group i, that responded to category k of item j.

$$g_c^i(jk) = \sum_{p=1}^{n_i} u_{cip} \delta_{ip}(jk) \tag{9}$$

By substituting Eqs.(2), (6) and (8) into each term of Eq.(3) and rearranging, The variation S_{T_c} in cluster c and the variation between groups S_{B_c} can be written respectively as:

$$S_{T_c} = \sum_{c=1}^{C} \sum_{i=1}^{K} \sum_{p=1}^{n_i} u_{cip} (Y_{cip} - \bar{Y}_c)^2$$

$$= \sum_{c=1}^{C} \sum_{j=1}^{R} \sum_{k=1}^{x_j} \sum_{l=1}^{R} \sum_{m=1}^{x_l} t_c(jk, lm) a_{cjk} a_{clm}$$

$$= \boldsymbol{a_c}^T T_c \boldsymbol{a_c} \tag{10}$$

$$t_c(jk, lm) = \{ f_c(jk, lm) - \frac{n_{cjk} n_{clm}}{n_c} \} \tag{11}$$

and

$$S_{B_c} = \sum_{c=1}^{C} \sum_{i=1}^{K} n_{ci} (\bar{Y}_{ci} - \bar{Y}_c)^2$$

$$= \sum_{c=1}^{C} \sum_{j=1}^{R} \sum_{k=1}^{x_j} \sum_{l=1}^{R} \sum_{m=1}^{x_j} b_c(jk, lm) a_{cjk} a_{clm}$$

$$= \boldsymbol{a_c}^T B_c \boldsymbol{a_c} \tag{12}$$

$$b_c(jk, lm) = \{\sum_{i=1}^{K} \frac{g_c^i(jk) g_c^i(lm)}{n_{ci}} - \frac{n_{cjk} n_{clm}}{n_c}\} \tag{13}$$

$f_c(jk, lm)$ in Eq.(11) is the sum of the memberships of individuals who responded to both the item j, category k and the item l, category m.

$$f_c(jk, lm) = \sum_{i=1}^{K} \sum_{p=1}^{n_i} u_{cip} \delta_{ip}(lm) \delta_{ip}(jl) \tag{14}$$

Hence, the objective function for maximizing the variations between-groups and partitioning into clusters can be given as follows:

$$\begin{aligned} \text{Ł} = \quad & \sum_{c=1}^{C} \sum_{i=1}^{K} \sum_{p=1}^{n_i} u_{cip} (\bar{Y_{ci}} - \bar{Y_c})^2 \\ & -\lambda_c (\sum_{c=1}^{C} \sum_{i=1}^{K} \sum_{p=1}^{n_i} u_{cip} (Y_{cip} - \bar{Y_c})^2 - 1) \\ & -\tau_0 \sum_{c=1}^{C} \sum_{i=1}^{K} \sum_{p=1}^{n_i} u_{cip} \log u_{cip} \\ & -\sum_{i=1}^{K} \sum_{p=1}^{n_i} \tau_{ip} (\sum_{c=1}^{C} u_{cip} - 1) \\ = \quad & \boldsymbol{a_c}^T B_c \boldsymbol{a_c} - \lambda_c (\boldsymbol{a_c}^T T_c \boldsymbol{a_c} - 1) \\ & -\tau_0 \sum_{c=1}^{C} \sum_{i=1}^{K} \sum_{p=1}^{n_i} u_{cip} \log u_{cip} \\ & -\sum_{i=1}^{K} \sum_{p=1}^{n_i} \tau_{ip} (\sum_{c=1}^{C} u_{cip} - 1) \end{aligned} \tag{15}$$

The third term of Eq.(15) is called entropy term, which serves as a regularizer [3] in Fuzzy c-Means Clustering. The coefficient τ_0 is for regulating the fuzziness. τ_{ip} is the Lagrangean multiplier.

From the optimality condition $\partial L / \partial \boldsymbol{a_c} = 0$, we have

$$(B_c - \lambda_c T_c) \boldsymbol{a_c} = \mathbf{0} \tag{16}$$

It is assumed that each individual responds to a single category within each item. Thus, for all j, i and p, following relation holds.

$$\sum_{k=1}^{x_j} \delta_{ip}(jk) = 1 \tag{17}$$

So, as convention we set

$$a_{cj1} = 0, \qquad c = 1, \cdots, C, \quad j = 1, \cdots, R \tag{18}$$

Thus, by eliminating the rows and columns corresponding to the first category in each item from the matrices B_c and T_c in Eq. (16), we solve Eq. (16) for each cluster. The solution a_{cjk} is the eigenvector corresponding to the largest eigenvalue λ_c in Eq.(16).

$$\lambda_c = \frac{\boldsymbol{a}_c^T B_c \boldsymbol{a}_c}{\boldsymbol{a}_c^T T_c \boldsymbol{a}_c} \tag{19}$$

In the first step, a_{cjk} is normalized such that

$$||\boldsymbol{a}_c|| = 1 \tag{20}$$

and then taking the volume of each cluster into account

$$\acute{a}_{cjk} = \sqrt{n_c} a_{cjk} \tag{21}$$

$$a^*_{cjk} = \acute{a}_{cjk} - \overline{Item}_{cj}, \tag{22}$$

$$\overline{Item}_{cj} = \frac{1}{n_c} \sum_{k=1}^{x_j} n_{cjk} \acute{a}_{cjk} \tag{23}$$

$\overline{Item}_{cj}$ is the mean value of $\acute{a}_{cjk}$ within item j in cluster c. Thus, we have

$$\bar{Y}_c = \frac{1}{n_c} \sum_{j=1}^{R} \sum_{k=1}^{x_j} n_{cjk} (a^*_{cjk} + \overline{Item}_{cj}) \tag{24}$$

$$\bar{Y}_{ci} = \frac{1}{n_{ci}} \sum_{j=1}^{R} \sum_{k=1}^{x_j} g_c^i(jk) (a^*_{cjk} + \overline{Item}_{cj}) \tag{25}$$

$$Y_{cip} = \sum_{j=1}^{R} \sum_{k=1}^{x_j} (a_{cjk} + \overline{Item}_{cj}) \delta_{ip}(jk) \tag{26}$$

From $\partial L / \partial u_{cip} = 0$, $\partial L / \partial \tau_{ip} = 0$

$$u_{cip} = \frac{\exp(A_{cip})}{\sum_{c'=1}^{C} \exp(A_{c'ip})} \tag{27}$$

where

$$A_{c'ip} = \frac{1}{\tau_0} (\bar{Y}_{c'i} - \bar{Y}_{c'})^2 - \frac{\lambda_{c'}}{\tau_0} (Y_{c'ip} - \bar{Y}_{c'})^2 \tag{28}$$

The solution to the problem, which maximize the objective function, is obtaind through the fixed point iteration like in the conventional Fuzzy c-Means clustering. Following is the algorithm.

(Step 1) Set τ_0, C and small positive number ε. Choose initial values of the membership u_{cip} randomly from unit interval [0, 1].

(Step 2) Solve eigenvalue problem (16) for each cluster and obtain a_{cjk}.

(Step 3) Normalize a_{cjk} by Eqs.(20) to (23).

(Step 4) Update membership u_{cip} by Eq.(27).

(Step 5) If the following termination condition is satisfied,

$$\max_{c,i,p}\{|u_{cip}^{NEW} - u_{cip}^{OLD}|\} < \varepsilon \tag{29}$$

then stop, otherwise go to (*Step* 2).

From the obtained a_{cjk}^*, each group mean is given as:

$$b_{ci} = \frac{1}{n_{ci}} \sum_{p=1}^{n_i} Y_{cip},$$

$$c = 1, \cdots, C, i = 1, \cdots, K \tag{30}$$

And, b_{ci} is also normalized as is done for a_{cjk}.

$$b_{ci}^* = b_{ci} - \overline{exY}_c, \tag{31}$$

$$\overline{exY}_c = \frac{1}{n_c} \sum_{i=1}^{K} \sum_{p=1}^{n_i} u_{cip} Y_{cip} \tag{32}$$

$\overline{exY}_c$ is the grand mean in each cluster. It is also necessary to normalize Y_{cip} such that

$$Y_{cip}^* = Y_{cip} - \overline{exY}_c \tag{33}$$

For the discriminant or classification of new individuals, the mean value b_{ip} of each group and the value Y_{ip} for each individual are given as the sum weighted by the membership u_{cip}.

$$\hat{b}_{ip} = \sum_{c=1}^{C} u_{cip} b_{ci}^* \tag{34}$$

$$\hat{Y}_{ip} = \sum_{c=1}^{C} u_{cip} Y_{cip}^* \tag{35}$$

Classifying process is as follows. The new individual p is classified into the group i whose $\hat{b}_{ip}$ is in the proximity to $\hat{Y}_{ip}$.

It should be noted that the membership of an individual to the fuzzy cluster can be calculated only when the group, to which the individual is belonging to, is known. When we are given the additional new data as in Table 2 and our task is to predict the group to which each of the individuals belongs, we

Table 2: Format of new data for prediction of group

individual	item 1			...	item R		
	1	...	x_1	...	1	...	x_R
1	$\surd$				$\surd$		
$\vdots$				$\vdots$			
N			$\surd$				$\surd$

have no idea to estimate the memberships to fuzzy clusters. Contrary to what was indicated for initially given data, we need to know the group to calculate the memberships. This point seems to be a major disadvantage against the conventional method in which no memberships are needed. Our sophisticated measure is to assume that each newly given individual belongs to every group one by one and then judge the group to which the individual is most likely to belong by comparing the values of $\hat{b}_{ip}$ and $\hat{Y}_{ip}$.

The membership of the data l is estimated as in Eqs.(27)-(28).

$$u_{cl} = \frac{\exp(A_{cl})}{\sum_{c'=1}^{C} \exp(A_{c'l})} \tag{36}$$

$$A_{c'l} = \frac{1}{\tau_0}(\bar{Y}_{c'i} - \bar{Y}_{c'})^2 - \frac{\lambda_{c'}}{\tau_0}(Y_{c'l} - \bar{Y}_{c'})^2 \tag{37}$$

where Y_{cl} is as in Eq.(26)

$$Y_{cl} = \sum_{j=1}^{R} \sum_{k=1}^{x_j} (a_{cjk} + \overline{Item}_{cj})\delta_l(jk) \tag{38}$$

Since the group i is unknown, by assuming that the individual belongs to one of each group, u_{cl} is estimated and then by using this value, like Eqs.(34)-(35),

$$\hat{b}_l = \sum_{c=1}^{C} u_{cl} b^*_{ci} \tag{39}$$

$$\hat{Y}_l = \sum_{c=1}^{C} u_{cl} Y^*_{cl} \tag{40}$$

are obtained. From the respective values of $\hat{b}_l$ and $\hat{Y}_l$, we can judge the group to which the individual is most likely to belong. The detail procedure of the judgment is exemplified by using the numerical example in the next section.

3 Numerical Example

A Questionnaire used in market research for opening a new restaurant is shown in Table 3 [4]. The purposes of the data analysis are as follows:

1. Estimate the ratio of latent customers who are not sure to become a customer of the restaurant.

2. Review the profiles of prospective customers and conduct an in-depth analysis for increasing the number of customers.

The summary of the responses about the intention of becoming a customer is as follows:

1. Yes:15

2. No:10

3. Not decided yet:25

Since half of the responders are not decided yet, it is most important to solicit and lure this group of customers. For the analysis of frequency count data from independent items, Cramer's V is used to examine the association or correlation between them [4]. The 15 dishes in the menu and their frequency count of binary outcome are used to examine the correlation with the intension of becoming a customer (Yes or No). We cross-tabulated the data to form 15 2-by-2 tables. The chi-square test statistic is a comparison of observed frequencies (O_j) to expected frequencies (E_j), where the observed frequencies come from the data and the expected frequencies are hypothetical occurrence of what one would encounter if the null hypothesis were true. Expected frequencies are calculated as the product of the values in the margins of the tables divided by the total sample size for both items. Pearson's ("uncorrected") chi-square test statistic is:

$$\chi^2 = \sum_j \frac{(O_j - E_j)^2}{E_j} \tag{41}$$

where O_j is the observed frequency in table cell j and E_j is the expected frequency in table cell j. The test statistic has $(r-1)(c-1)$ degree of freedom, where r represents the number of rows in the frequency table and c represents its number of columns. Cramer's V is calculated by correcting the chi squared statistic for sample and table size:

$$V = \sqrt{\frac{\chi^2}{n(k-1)}} \tag{42}$$

where n is the total number (frequency) and k is either r or c (whichever is smaller). On the basis of the Cramer's V, we selected 6 items from those in Question 2, that affect the customers' intention. The 15 responders who answered affirmatively and 10 negatively are classified by our proposed method

for testing the classification performance. All responders are correctly classified. By the conventional quantifying method, the classification ratio is 80 percent. The nonlinearity due to the membership function seems to have contributed to this improvement. The weighting coefficient for fuzzification in clustering provides the gradation of nonlinearity. The totally fuzzy clusters ($\tau_0 = \infty$) derive the same results as one by the conventional method, since all individuals have the same memberships.
Now we illustrate how to classify the newly given individual whose membership in a group is not known, hence the membership in a cluster is also unknown. We predicted how many of them among 25 (50 percent) responders, who did not give definite answer(Yes or No), have the similar characteristics as those who are willing to become a customer. A part of the prediction results is given in Table 4.

Table 4 shows the way to classify group unknown individuals. The left most column indicates the final judgment of group. Individual's number is shown in Ind No. column. The real number shows the absolute difference from the group mean whose value $\hat{b}_l$ is in the proximity to the value $\hat{Y}_l$. The number 2 and 1 indicate the position of group means. For example, Individual 1 is classified as Group 1(Expected customer) when the individual is assumed to be in Group 1 (as shown in the upper row) and also when assumed to be in Group 2 (as shown in the lower row), the rational judgment is apparently Group 1 since $\hat{Y}_l$ appears in the right side of the Group 1 mean which is denoted by 1. So the final judgment is Group 1. As to individual 10, in both cases (upper and lower row) the real values are written between the two group means and the judgments are different from each other. In this ambiguous case, the smaller distance is chosen. So the final judgment is Group 1. Most of the other individuals are apparently classified. Among the 25 responders, 17 of them have the similar characteristics with the positive responders and they can be considered to be the expected customers. The ratio of latent customers is 64 percent.

4 Conclusion

We have proposed a nonlinear version of the classical discriminant analysis technique known as the second Hayashi's quantification method for categorical data. Although the notion of membership and FCM clustering algorithm are introduced, the aim of the proposed algorithm is not only the analysis of locally partitioned data but to make the underlying discriminant model nonlinear. The fuzzy clusters are implicitly used for the analysis and the classification of new data.

References

[1] C. Hayashi, On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematical statistical point

of view, Ann. the Institute of Statistical Math., Vol.3, pp.69-98, 1952.

[2] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.

[3] S. Miyamoto and M. Mukaidono, Fuzzy c-means as a regularization and maximum entropy approach, Proc. IFSA'97, Vol.II, pp.86-92, 1997.

[4] T. Kan, Analysis of Questionnaire, Gendaisugakusha,1998(in Japanese).

Table 3: Questionnaire sheet

Our new restaurant ABC will start to service
on November 1. Taking our menus into account,
please answer each questions below.

Question 1 Do you have an intention to be
a customer of restaurant ABC?

1. Yes 2. No 3. Not decided yet

Question 2 What lunch do you typically take?
Please select up to 5 dishes.

1. Cutlet bowl 2. Tempura bowl 3. Eel bowl
4. Beaf bowl 5. China bowl 6. Sashimi lunch
7. Broiled fish lunch 8. Oden lunch
9. Yakitori lunch 10. Pork lunch
11. Hamburger lunch 12. Larmen lunch
13. Soba lunch 14. Rice and curry
15. Sushi

Question 3 Which one do you think
should come along with your meal?

	prefer	neutral	dislike
A. Liquor ...	1	2	3
B. Sweets ...	1	2	3
C. Coffee ...	1	2	3

Question 4 Gender

1. Male 2. Female

Question 5 Age

☐ years

Table 4: Prediction of latent customers

Judge	Ind No.	Distance from group mean 2	1	Gr No.
1	1		0.53	1
			0.03	1
2	2		0.72	2
			0.72	2
2	3	0.55		2
		0.55		2
1	10		0.45	1
			0.82	2
2	11	0.46		2
		0.46		2
1	12			0.79	1
				0.79	1