International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:3, No:6, 2009

A Formulation of the Latent Class Vector Model for Pairwise Data

Tomoya Okubo, Kuninori Nakamura, and Shin-ichi Mayekawa

Abstract**—**In this research, a latent class vector model for pairwise data is formulated. As compared to the basic vector model, this model yields consistent estimates of the parameters since the number of parameters to be estimated does not increase with the number of subjects. The result of the analysis reveals that the model was stable and could classify each subject to the latent classes representing the typical scales used by these subjects.

Keywords**—**finite mixture models, latent class analysis, Thrustone's paired comparison method, vector model

I. Introduction

THURSTONE's paired comparison method [1], which projects stimuli into a uni-dimensional scale based on preference (choice) data has been highly influential not only in the field of psychometrics but also in related domains such as marketing research. In this method, subjects are asked to compare two stimuli (i, j) and indicate their preferences for some pairs. After the paired comparison, we obtained the frequency with which i (or j) is preferred to the other for each paired comparison. Then under some assumptions, we are able to place stimuli on a uni-dimensional scale based on the data.

Tsai and Böckenholt [2] pointed out the following three advantages in collecting data using the paired comparison method. First, the method imposes minimal constraints on the response behavior. Second, internal consistency checks are available, which allow for the identification of judges who are systematically inconsistent in their judgments. Third, paired comparison data provide a rich source of information about individual differences and similarity relationships in the evaluation of preference data.

Some plausible models have been presented to describe individual differences in choice data ([3]-[10]). Böckenholt and Böckenholt [11] presented framework of a latent class vector model (LC vector model) and a latent class ideal-point model (LC ideal-point model) of discrete choice data; in these models, fewer parameters have to be estimated as compared with the basic vector model. The model simultaneously clusters the subjects into a small number of latent classes and constructs a configuration of the stimuli. However, they did not formulate the LC vector model for the pairwise data.

The purpose of the present article is to formulate the LC vector model to the pairwise data in order to describe individual differences with a fewer number of parameters. Also, we analyze real data in order to confirm utility of the model.

II. Latent Class Vector Model for Pairwise Data

This model is a latent class extension ([12], [13]) of the vector model in which the subjects are treated as observations to be classified in paired comparison data.

Let n_{ijk} be the number of comparisons between i and j by subject k, and let f_{ijk} denote the frequencies with which i is preferred to j by subject k.

We assume that the frequencies f_{ijk} are distributed as

$$f_{ijk} = \mathrm{Bin}(n_{ijk}, \pi_{ijq}) \tag{1}$$

where

$$\pi_{ijq} \approx \psi_{ijq} = 1 / \exp(-1.7 z_{ijq}) \tag{2}$$

and

$$z_{ijq} = \xi_{iq} - \xi_{jq}, \quad \mathbf{z}_q = \mathbf{G}\boldsymbol{\xi}_q = \mathbf{G}\mathbf{X}\boldsymbol{\beta}_q \tag{3}$$

where $\mathbf{X}$ is a principle component score matrix, and $\boldsymbol{\beta}$ is a principle component loading matrix.

Also, $\mathbf{G}$ is a design matrix, which is described as

T. Okubo is with The National Center for University Entrance Examinations, Meguro-ku, Tokyo 153-8501 Japan (e-mail: okubo@rd.dnc.ac.jp).

K. Nakamura is with Graduate School of Decision Science and Technology, Tokyo Institute of Technology, Meguro-ku Tokyo 152-8550 Japan.

S. Mayekawa is with Graduate School of Decision Science and Technology, Tokyo Institute of Technology, Meguro-ku Tokyo 152-8550 Japan.

$$\mathbf{G} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \quad (4)$$

In this formula, $\boldsymbol{\xi}_{iq}$ denotes the scale value for a stimulus i by the subject, who belongs to the class q.

The objective function to be maximized is

$$L = \prod_{k=1}^{K}\prod_{q=1}^{Q}\left(\prod_{i>j}\pi_{ijq}^{\ f_{ijk}}(1-\pi_{ijq}^{\ (n_{ijk}-f_{ijk})})\right)^{d_{kq}} \quad (5)$$

with constraints

$$\mathbf{1x}' = \mathbf{0}',\ \mathbf{x}'\mathbf{x} = \alpha\mathbf{I},\ \boldsymbol{\beta}'\boldsymbol{\beta} = \text{diagonal} \quad (6)$$

where d_{kq} is an unobserved missing classification data, which indicates subject k belongs to class q or not.

$$d_{kq} = \begin{cases} 1 & \text{if subject } k \text{ belongs to class } q \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

To maximize a such a function, we employ EM algorithm [14].In the E step, we compute the conditional expectation of d_{kq}, which is written as

$$E(d_{kq} \mid \mathbf{f}_k) = \frac{\prod_{i>j}\pi_{ijq}^{\ f_{ijk}}(1-\pi_{ijq})^{(n_{ijk}-f_{ijk})}\rho_q}{\sum_{q=1}^{Q}\prod_{i>j}\pi_{ijq}^{\ f_{ijk}}(1-\pi_{ijq})^{(n_{ijk}-f_{ijk})}\rho_q} = h_{kq} \quad (8)$$

In the M step, following the complete-data log likelihood is maximized

$$\begin{aligned} E\ln L = &\sum_{q=1}^{Q}\sum_{i>j}\left(\sum_{k=1}^{K}h_{kq}f_{ijk}\right)\log\left(\pi_{ijq}\right) \\ &+\sum_{q=1}^{Q}\sum_{i>j}\left(\sum_{k=1}^{K}h_{kq}(n_{ijk}-f_{ijk})\right)\log\left(1-\pi_{ijq}\right) \\ &+\sum_{q=1}^{Q}\left(\sum_{k=1}^{K}h_{kq}\right)\log(\rho_q) \end{aligned} \quad (9)$$

Further, the mixing proportion for the class q is computed as

$$\rho_q = \frac{\sum_{k=1}^{K}h_{kq}}{\sum_{q=1}^{Q}\sum_{k=1}^{K}h_{kq}} \quad (10)$$

III. Real Data Analysis

A. Data

We applied the LC vector model for pairwise data to job preference data. A total of 79 undergraduates and graduates served as volunteer subjects for this study (mean age=21.0 years). Using the paired comparison designs, the subjects were asked to identify the better job from a pair of jobs. The following 10 jobs were compared: politician (Po), programmer (Pg), medical doctor (Md), artist (Ar), salesperson (Sa), banker (Bk), journalist (Jo), researcher (Re), civil servant (Cs), bus driver (Bd).

In this analysis, the following three points are assumed.

(1) Every single judgment is mutually independent.

(2) The image concerning a job is common across subjects.

(3) The individual differences in preference are described as differences in the weight of the job image. Note that this analysis extracts data on subjects whose job image is the same as that in the preliminary research.

B. Results

The LC vector model for pairwise data was applied to the data shown earlier. The results of the analysis are presented in Tables I, II, and III and in Fig. 2. Further, for a comparative evaluation of the LC and basic vector models, the latter model was applied to the data (Fig. 1). In this case, r (the number of dimensions) was set to two beforehand in order to draw a biplot. Table I contains the principal component score matrix $\mathbf{X}$ estimated by LC vector model and Table II presents the principle component loading matrix $\boldsymbol{\beta}_q$ and mixing proportion $\boldsymbol{\rho}_q$ estimated using the LC vector model. In addition, Table III contains the scale values $\boldsymbol{\xi}$ for each class. Fig. 1 and 2 illustrate the biplot of the model.

TABLE I
ESTIMATED $\mathbf{X}$ USING THE LC VECTOR MODEL

Job	$\mathbf{x}_1$	$\mathbf{x}_2$
Politician (Po)	0.793	-1.419
Programmer (Pg)	-0.315	0.039
Medical doctor (Md)	2.266	0.132
Artist (Ar)	-1.161	1.426
Salesperson (Sa)	-0.758	-0.464
Banker (Bk)	0.434	-0.368

Journalist (Jo)	-0.426	1.039
Researcher (Re)	0.286	0.539
Civil servant (Cs)	0.193	0.915
Bus driver (Bd)	-1.313	-1.838

TABLE II
ESTIMATED $\boldsymbol{\beta}$ AND $\boldsymbol{\rho}$ USING THE LC VECTOR MODEL

Class	$\boldsymbol{\beta}_1$	$\boldsymbol{\beta}_2$	$\boldsymbol{\rho}_q$
Class 1	1.057	0.219	0.085
Class 2	0.329	0.723	0.208
Class 3	0.381	0.307	0.267
Class 4	-0.183	0.696	0.153
Class 5	-0.171	0.216	0.159
Class 6	-0.643	0.656	0.128

In the LC vector model, the number of latent classes is determined based on the Akaike information criterion (AIC). In this analysis, the AIC was the lowest when the number of classes was 6. The number of vectors in Fig. 2 is less than that in Fig. 1, because we classified the 79 vectors (subjects) into 6 vectors (classes). As is evident in these figures, the configurations between the basic and LC vector models are almost identical. Further, as is clear in Table III, when a subject from Class 1 compares a medical doctor and a journalist, the probability of selecting the medical doctor over the journalist is higher. This is because in Class 1, the scale value for the medical doctor is 2.424 whereas that for the journalist is -0.223. On the other hand, when a subject from Class 4 compares the medical doctor and the journalist, the probability of selecting the journalist over the medical doctor is higher. This is because in Class 4, the scale value for the medical doctor in Class 4 is -0.323 while that for the journalist is 0.801.

TABLE III
ESTIMATED $\boldsymbol{\xi}$ USING THE LC VECTOR MODEL

Job	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
Politician	0.527	-0.765	-0.134	-1.133	-0.442	-1.441
Programmer	-0.324	-0.075	-0.108	0.085	0.062	0.228
Medical doctor	2.424	0.841	0.904	-0.323	-0.359	-1.370
Artist	-0.915	0.649	-0.005	1.205	0.507	1.682
Salesperson	-0.903	-0.585	-0.431	-0.184	0.029	0.183
Banker	0.378	-0.123	0.052	-0.336	-0.154	-0.520
Journalist	-0.223	0.611	0.157	0.801	0.297	0.956
Researcher	0.420	0.484	0.274	0.323	0.068	0.170
Civil servant	0.404	0.725	0.354	0.602	0.165	0.476
Bus driver	-1.790	-1.761	-1.065	-1.039	-0.172	-0.361

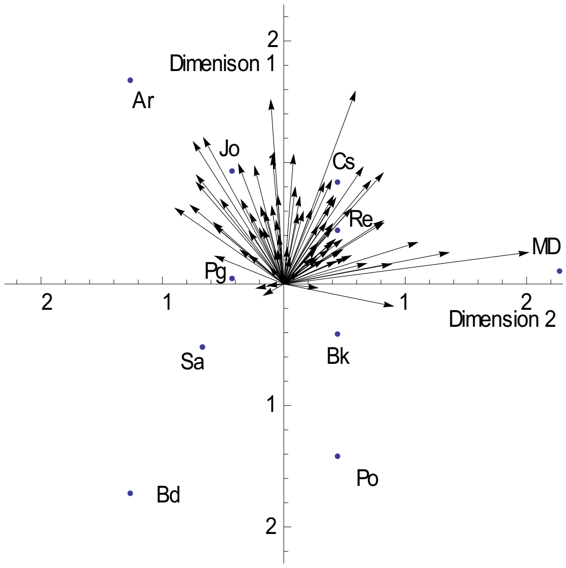

Fig. 1 Biplot of the vector model

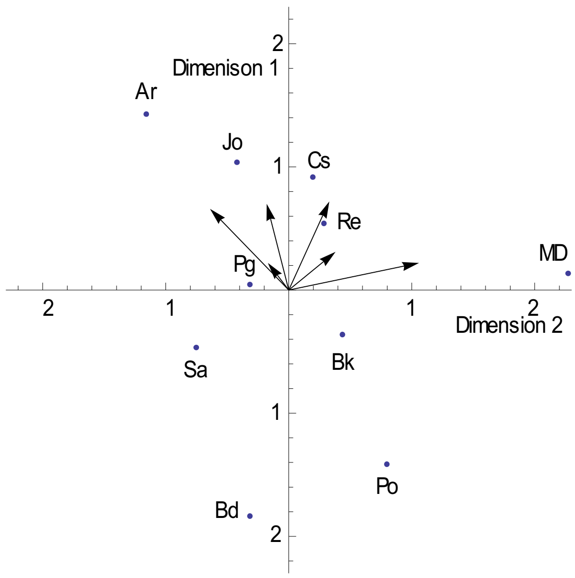

Fig. 2 Biplot of the LC vector model

IV. Conclusion

In this article, we applied the LC vector model to pairwise data. This enables the analysis of the individual differences in pairwise data with fewer parameters than in the basic vector model. Furthermore, the utility of the model was assessed by analyzing data on job selection. We classified the subjects into certain classes by using this model. We obtain a sensitive preference scale with the pairwise data, which cannot be obtained with rating data or discrete data. Moreover, by analyzing the data with the LC vector model, which was presented in this article, we can simultaneously represent the characteristics of the stimuli and the features of the typical choice of subjects. Furthermore, it is easy to interpret the parameters in this model. However, the paired comparison method increases the number of judgments. It is necessary to expand the model in order that the missing data can be treated.

References

[1] J. Arbuckle and J. H. Nugent, "A general procedure for parameter estimation for the law of comparative judgement," *British Journal of Mathematical and Statistical Psychology*, vol. 26, 1973, pp. 240-260.

[2] P. Slater, "The analysis of personal preferences," *British Journal of Statistical Psychology*, 13, 1960, pp. 119-135.

[3] L. R. Tucker, "Intra-individual and Inter-individual multidimensionality", in *Psychological scaling*: *Theory and Applications*, H. Gulliksen, and S. Messick Eds, New York: Wiley, 1960, pp. 155-167.

[4] J. F. Bennett and W. L. Hays, "Multidimensional unfolding: determining the dimensionality of ranked preference data," *Psychometrika*, vol. 25, 1960, pp. 27-43.

[5] J. D. Carroll, Individual differences and multidimensional scaling. in *Multidimensional scaling*, R. N. Shepard, A. K. Romney, and S. B. Nerlove Eds, New York: Seminar Press, 1972, pp. 105-155.

[6] Y. Takane, "Maximum likelihood estimation in the generalized case of Thurstone's model of comparativejudgment," *Japanese Psychological Research*, 22, 1980, pp. 188-196.

[7] Y. Takane, "Analysis of covariance structures and probabilistic binary choice data," in *New developments in psychological choice modeling advances in psychology*, G. DeSoete, H. Feger, and K. C. Klauer Eds, Amsterdam: Elsevier Science Ltd, 1989, pp. 139-160.

[8] Y. Takane, "Choice model analysis of the ``pick any/n" type of binary data," *Japanese Psychological Research*, 40, 1998, pp. 31-39.

[9] U. Böckenholt, and I. Böckenholt, "Constrained latent class analysis: Simultaneous classification and scaling of discrete choice data," *Psychometrika*, 56, 1991, pp. 699-716.

[10] L. A. Goodman, "Exploratory latent structure analysis using both identifiable and unidentifiable models," *Biometrika*, 61, 1974, pp. 215-231.

[11] P. F. Lazarsfeld, and N. W. Henry, *Latent structure analysis*. Boston: Houghton Mifflin Company, 1968.

[12] A. P. Dempster, N. M. Laired, and D. B. Rubin, "Maximum likelihood from incomplete sata via the EM algorithm," *Journal of Royal Statistical Society*, Series B, 39, 1977, pp. 1-38.

[13] G. DeSoete, and S. Winsberg, "A latent class vector model for preference ratings," *Journal of Classification*, 10, 1993, pp. 195-218.

[14] U. Böckenholt, "A Thurstonian analysis of preference change," *Journal of Mathematical Psychology*, 46, 2002, pp. 300-314.